(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36685 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US00/31809

(22) International Filing Date:
17 November 2000 (17.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/166,056	17 November 1999 (17.11.1999)	US
60/166,106	17 November 1999 (17.11.1999)	US

(71) Applicant *(for all designated States except US)*: **NYXIS NEUROTHERAPIES, INC.** [US/US]; 430 West Deming, 2nd Floor, Chicago, IL 60614 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KROES, Roger, A.** [US/US]; 25174 N. Virginia, Lake Zurich, IL 60047 (US). **MOSKAL, Joseph, R.** [US/US]; 515 W. Roscoe, Chicago, IL 60657 (US). **YAMAMOTO, Hirotaka** [US/US]; 1810 George Court, Glenview, IL 60025 (US).

(74) Agent: **HALLORAN, Patrick, J.**; McDonnell Boehnen Hulbert & Berghoff, Suite 3200, 300 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/36685 A2

(54) Title: DIFFERENTIAL GENE EXPRESSION IN CANCER

(57) Abstract: The invention is directed towards methods for ascertaining gene expression characteristic for cancer, in particular brain cancers such as glioblastoma, and the sequences identified thereby. Compositions, methods and kits encompassing such are provided herein.

DIFFERENTIAL GENE EXPRESSION IN CANCER

Field of the Invention

The invention relates to the field of cancer, in particular characteristic genes and gene expression useful in screening for, diagnosis of, monitoring of, and therapeutic treatment of cancer. Further, the invention relates to age-related differential expression of genes in cancer.

Background of the Invention

Cancer can develop in any tissue of any organ at any age. Most cancers detected at an early stage are potentially curable; thus, physicians need a heightened awareness of predisposing inherited and environmental factors. The ability to screen patients for genetic predisposition for cancer can greatly assist in the monitoring of high-risk patients for early signs of cancer, and thus allowing for early intervention. (See for example, The Merck Manual of Diagnosis and Therapy, 16th ed., Merck & Co., (1992)).

Malignant brain tumors (for example glioma, meningiomas, and schwannomas) are common, with an incidence of 4.5 per 100,000. The most common tumor types in adults are gliomas and meningiomas. The most common tumors in children are astrocytomas, medulloblastomas, ependymomas, and brain stem gliomas. In children, brain tumors are one of the most common causes of death from cancer. (See for example, Professional Guide to Disease, 3rd ed., Springhouse Corp., (1989)).

Clinically, brain tumors can be characterized by their cell type and location, along with other phenotypic clues. Malignant brain tumors are sometimes catagorized as glioblastoma multiforme (spongioblastoma multiforme), astrocytoma, oligodendroglioma, ependyoma, medulloblastoma, meningioma, schwannoma, and pituitary tumors. It is also possible that cancer originating in other tissues, such as lung, liver, pancreas, colon, prostate etc., can metastasize to the brain, thus forming tumors that are not of brain origin, potentially causing confusion as to the source of cancer.

Cancer is a cellular malignancy whose unique trait - loss of normal control mechanisms - results in unregulated growth, lack of differentiation, and ability to invade local tissues and metastasize. Thus cancer cells are unlike normal cells, and are potentially identifiable by not only their phenotypic traits, but also by their biochemical and molecular biological characteristics. In particular, the altered phenotype of cancer cells indicates altered gene activity, either unusual

gene expression, or gene regulation. Identification of gene expression products or proteins associated with cancer cells will allow for the molecular characterization of malignancies. The ability to specifically characterize suspected cancers, and to potentially identify not only cell type, but also predisposition for metastasis and any sensitivity to particular anti-cancer therapy, is most useful for determining not only the course of treatment, but also the likelihood of success.

Thus, the discovery of specific, brain tumor characteristic gene expression is a useful and important tool useful in screening for, diagnosis of, monitoring of, and therapeutic treatment of brain cancer. In particular, provided herein are methodologies and sequences that are differentially expressed in cancer from age-differentiated patients.

Summary of the Invention

The identification of characteristic, nucleic acid signals is a useful and important discovery which allows for compositions, assays, kits and reagents suitable for the characterization of various brain cancers. Provided herein are reagents and methods for ascertaining the propensity of a cell for malignant phenotype said cell being isolated or in a biological sample, said method comprising assaying a cell or biological sample to be tested for a signal indicating the transcription of a nucleic acid transcript. In a preferred embodiment, the nucleic acids are substantially identical to the sequences of SEQ ID NOS. 1-184, or fragments thereof. Also provided are methods for monitoring cancer progession or the effectiveness of a treatment regimen, and methods for identifying compounds that affect expression of genes involved in cancer.

One of ordinary skill in the art will be able to understand and ascertain modifications and embodiments of the present invention that fall within the spirit and scope of the disclosure as described below.

Brief Description of the Figures

Figure 1. Relationship between patient age at diagnosis and glioma survival

The survival pattern for Grade IV astrocytoma (GBM) patients according to four age strata is illustrated. The apparent differences between the <35 and the 35-50 year group are not statitically significant, but the survival for the <50 year group as a whole was statistically different from the 50-65 and the >65 year groups (Wilcoxon test, p=.002).

Figure 2. Northern analysis of hsp60 mRNA. 25 μg of total RNA was isolated, electrophoresed through 1.2% agarose-formaldehyde gels, transferred to nylon membranes and hybridized with a uniformly (^{32}P)-labeled hsp60-specific cDNA probe. (A): hsp60 expression in normal (NL) brain and GBMs. Patient age at diagnosis is depicted. (B): Developmental expression of hsp60 in normal brain tissue.

Figure 3. Normal Developmental Expression of Heat Shock Proteins in Human Brain. 25 g of total RNA was isolated, electrophoresed through 1.2% agarose-formaldehyde gels, transferred to nylon membranes and hybridized with a uniformly (^{32}P)-labeled cDNA probes specific for hsp27, hsp70, hsc72, hsp90α, hsp90β, and GRP78.

Figure 4. Differential Expression of Heat Shock Proteins in Human Gliomas 25 μg of total RNA was isolated, electrophoresed through 1.2% agarose-formaldehyde gels, transferred to nylon membranes and hybridized with a uniformly (^{32}P)-labeled cDNA probes specific for hsp27, hsp70, hsc72, and hsp90β.

Detailed Description of the Invention

It is believed that brain tumorigenesis results from complex interactions of multiple and cumulative genetic alterations. These events lead to either the activation of various oncogenes, overriding regulatory signals which control cell proliferation, or inactivation of tumor suppressor genes, resulting in the uncontrolled growth of cells. (*See for example* Burck et al., Oncogenes, Springer-Verlag, New York, 1988). The identification and characterization of subsets of the genes associated with such uncontrolled growth is essential in order to understand the process of malignancy, but more importantly, useful for the identification of specific cancerous tissues, and tissues that are premalignant, and potentially predisposed for it.

Cancer is defined herein as any cellular malignancy for which a loss of normal cellular controls results in unregulated growth, lack of differentiation, and increased ability to invade local tissues and metastasize. Cancer may develop in any tissue of any organ at any age. Cancer may be an inherited disorder or caused by environmental factors or infectious agents; it may also result from a combination of these.

The differential expression of genes that regulate cell growth, migration, and other functions enables a cell to grow out of control and become cancerous. In many cases, the

activation of oncogenes, which override the intrinsic cellular growth regulatory commands of a cell, as well as the inactivation of tumor suppressor genes, which normally hold tumor formation in check, renders tumor cells free of growth restraints. The identification and characterization of these differentially expressed genes in malignant tumors will facilitate the understanding of the basic nature of the malignancy and yield novel molecular markers useful in diagnosis and treatment. For the purposes of utilizing the present invention, the term cancer includes both neoplasms and premalignant cells.

In one embodiment, the present invention is useful for the diagnosis and treatment of many types of cancers including, for example, cancers of the breast, prostate, colon, and lung. In a preferred embodiment, the reagents and methodologies provided herein are useful for the diagnosis and treatment of brain cancer. Brain tumors (or brain cancer) arise as a result of complex interactions of multiple and cumulative genetic alterations. Brain cancer is defined herein as any cancer involving a cell of neural origin. Examples of brain cancers include but are not limited to intracranial neoplasms such as those of the skull (i.e., osteoma, hemangioma, granuloma, xanthoma, osteitis deformans), the meninges (i.e., meningioma, sarcoma, gliomatosis), the cranial nerves (i.e., glioma of the optic nerve, schwannoma), the neuroglia (i.e., gliomas) and ependyma (i.e., ependymomas), the pituitary or pineal body (i.e., pituitary adenoma, pinealoma), and those of congenital origin (i.e., craniopharygioma, chordoma, germinoma, teratoma, dermoid cyst, angioma, hemangioblastoma) as well as those of metastatic origin.

As demonstrated herein, it has been discovered that brain cancer cells, in particular glioma cells, express certain nucleic acid sequences at a higher level than that found in normal brain cells, for example fetal astrocytes. Similarly, it has been found that this expression is most commonly detected as a nucleic acid, usually mRNA which is expressed from an activated gene, resulting in a detectable nucleic acid signal corresponding to the transcript from a gene. The present invention teaches a specific array of gene signals, i.e. expressed genes, mRNA transcripts, which indicate a cells propensity for a malignant phenotype in cancer. In a preferred embodiment, the gene sequences provided herein are indicative of brain cancer. In addition, the present invention provides an assay system for the detection of cancer and the monitoring of treatment progress. In one embodiment, a panel comprising one or more of SEQ ID NOS. 1-141, or fragments or complements thereof, may be utilized to identify cancerous cells. In a preferred

embodiment, the panel comprises one or more of SEQ ID NOS. 68, 69 or 183, or fragments or complements thereof.

One of the most significant factors impacting the survival of patients with glioblastomas (GBM) is age at primary diagnosis. Patients diagnosed prior to the age of 50 years survive significantly longer than those diagnosed after the age of 50, with median survival of 24 months and 8 months, respectively. This difference in survival is independent of performance status and appears to be unrelated to treatment. The cellular mechanisms for this age/prognosis correlation are not known. Several age-related chromosomal aberrations in GBM have been recently described, and include +7, amplifications on 7, -18q and –10 in tumors from older patients. Additionally, +17q, -Xp, -5q, and –10q have been found to occur in tumors from younger patients. These data strongly suggests a molecular basis for this poor patient survival. Provded herein is a DDRT-PCR based approach to define molecular changes associated with this age-dependent survival of GBM patients, and a panel of differentially expressed genes from tumors resected from these disparate patient populations. The present invention further provides novel nucleic acid sequences representing genes and the polypeptides encoded thereby that are involved in cancer progression.

In one embodiment, the expression of a panel of sequences comprising one or more of SEQ ID NOS. 142-182, or fragments or complements thereof, may be assayed to characterize the tumors of old vs. young patients. In a preferred embodiment, the panel comprises one or more of SEQ ID NOS. 142-174, or fragments of complements thereof, where over-expression in tumors of old patients as compared to young patients of the sequences is detected. In a more preferred embodiment, the panel comprises one or more of SEQ ID NOS. 142, 143, 144, 147, 149, 162 or 173, or fragments or complements thereof, where increased expression of the sequences in tumors of old patients as compared to young patients is detected. In another preferred embodiment, the panel comprises one or more of SEQ ID NOS. 175-182, or fragments of complements thereof, where decreased expression of the sequences in tumors of old patients as compared to young patients is detected.

I. General Methodology

Within this application, unless otherwise stated, the techniques utilized may be found in any of several well-known references including: *Molecular Cloning: A Laboratory Manual*

(Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, CA), Berger et al., Guide to Molecular Cloning Techniques, *Methods in Enzymology*, Vol. 152, Academic Press, Inc., (1987); Davis et al., *Basic Methods in Molecular Biology*, Elsevier Science Publishing Co., Inc. (1986); Ausubel et al., *Short Protocols in Molecular Biology*, 2nd ed., John Wiley & Sons, (1992), Grinsted et al., *Plasmid Technology, Methods in Microbiology*, Vol. 21, Academic Press, Inc., (1988); Symonds et al., *Phage Mu*, Cold Spring Harbor Laboratory Press (1987), Guthrie et al., Guide to Yeast Genetics and Molecular Biology, *Methods in Enzymology*, Vol. 194, Academic Press, Inc., (1991), *PCR Protocols: A Guide to Methods and Applications* (Innis, et al. 1990. Academic Press, San Diego, CA), McPherson et al., *PCR Volume 1*, Oxford University Press, (1991), *Culture of Animal Cells: A Manual of Basic Technique, 2nd Ed.* (R.I. Freshney. 1987. Liss, Inc. New York, NY), and *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E.J. Murray, The Humana Press Inc., Clifton, N.J.). The basic principles of eukaryotic gene structure and expression are generally known in the art. (*See for example* Hawkins, *Gene Structure and Expression*, Cambridge University Press, Cambridge, UK, 1985; Alberts et al., *The Molecular Biology of the Cell*, Garland Press, New York, 1983; Goeddel, *Gene Expression Technology, Methods in Enzymology*, Vol. 185, Academic Press, Inc., (1991); Lewin, *Genes VI*, Oxford Press, Oxford, UK, 1998). Each of the above-mentioned references and any of those listed below including issued patents are hereby incorporated by reference.

For the purposes of this application, certain terms are defined below. The meaning of these terms are generally understood by those of skill in the art, and the descriptions provided herein are provided merely as additional guidance.

A *transcriptional regulatory region* is defined as any region of a gene involved in regulating transcription of a gene, including but not limited to promoters, enhancers and repressors. A *transcriptional regulatory element* is defined as any element involved in regulating transcription of a gene, including but not limited to promoters, enhancers and repressors. A *promoter* is a regulatory sequence of DNA that is involved in the binding of RNA polymerase to initiate transcription of a gene. A *gene* is a segment of DNA involved in producing a peptide, polypeptide or protein, including the coding region, non-coding regions preceding ("leader") and following ("trailer") the coding region, as well as intervening non-coding sequences ("introns") between individual coding segments ("exons"). *Coding* refers to the representation by the nucleic

acid of amino acids, start and stop signals in a three base "triplet" code. Promoters are often upstream ("5' to") the transcription initiation site of the corresponding gene. Other regulatory sequences of DNA in addition to promoters are known, including sequences involved with the binding of transcription factors, including response elements that are the DNA sequences bound by inducible factors. *Enhancers* comprise yet another group of regulatory sequences of DNA that can increase the utilization of promoters, and can function in either orientation (5'-3' or 3'-5') and in any location (upstream or downstream) relative to the promoter. Preferably, the regulatory sequence has a positive activity, i.e., binding of an endogeneous ligand (e.g. a transcription factor) to the regulatory sequence increases transcription, thereby resulting in increased expression of the corresponding target gene. The term *operably linked* refers to the combination of a first nucleic acid fragment representing a transcriptional control region having activity in a cell joined to a second nucleic acid fragment encoding a reporter or effector gene such that expression of said reporter or effector gene is influenced by the presence of said transcriptional control region.

A *polypeptide* refers to an amino acid sequence encoded by a nucleic acid, a fragment thereof, or a nucleic acid comprising a nucleic acid of this invention. Preferably, the nucleic acids of this invention are selected from those described by SEQ ID NOS. 1-184.

A *nucleic acid* or *protein fragment* relates to a portion of a larger sequence from which the fragment is derived, where the fragment is useful for performing the methods described herein. For instance, a particular sequence described within this application may contain irrelevant nucleotides derived from a cloning vector or primer used in amplifying the nucleic acid (i.e., HindIII site, poly-A, poly-T). Those nucleotides could be deleted from the particular sequence, resulting in a functional fragment of the larger sequence. Similarly, a portion of a sequence (i.e., 15 nucleotides of a 200 bp nucleic acid) may be utilized for detecting expression of a gene sequence within a cell. A protein fragment is a sequence of amino acids derived from a protein that is functional, as an immunogen, a probe to detect autoantibodies, or to identify relevant ligands, for example.

A *responsive element* is a portion of a transcriptional control region that induces expression of a nucleotide sequence following the interaction of a cell with a compound. There may be multiple responsive elements within a single transcriptional control region and each of these elements may function independently of any other elements of that transcriptional control

region. Thus, a responsive element may be incorporated into a reporter gene vector independent from the remainder of the transcriptional control region from which it is derived and function to drive expression of the reporter gene under the proper conditions.

The terms *overexpressed* or *underexpressed* typically relate to expression of a nucleic acid sequence or protein in a tumor cell at a higher or lower level, respectively, than that level typically observed in a non-tumor cell (i.e., normal control). For instance, a particular sequence may be over- or under-expressed in cells or tissue obtained from a patient older than 60 years ("old" patient) as compared to a sample of cells or tissue obtained from a patient younger than 45 years old ("young" patient). In certain cases, the terms *overexpressed* or *underexpressed* may also relate to the expression level in a cell that has been contacted by a compound and compared to the expression level in a similar cell that has not been contacted by the compound.

The terms *cancer cell* and *tumor cell* and the like may be used interchangeably and relate to cells found within a cancerous growth or tumor. The reagents and methodologies provided herein are applicable to the detection, diagnosis, and treatment of many types of cancers. In a preferred embodiment, the reagents and methodologies provided herein are useful for the detection, diagnosis, and treatment of brain cancer.

For the purposes of this application, hybridization is typically performed under stringent conditions. The term *stringent conditions* refers to hybridization and washing under conditions that permit only binding of a nucleic acid molecule such as an oligonucleotide or cDNA molecule probe to highly homologous sequences. For example, a stringent wash solution is 0.015 M NaCl, 0.005 M NaCitrate, and 0.1% SDS used at a temperature of 55°C-65°C. Another stringent wash solution is 0.2X SSC and 0.1% SDS used at a temperature of between 50°C-65° C.

A nucleic acid, DNA, RNA or amino acid sequence is *identical* or the *same* as another sequence where the sequences are identical. A nucleic acid, DNA, RNA or amino acid sequence is *substantially identical* or *substantially the same* as another sequence where the sequences are 50-100% identical. In a preferred embodiment, substantially identical sequences share 60-100% identity, more preferably 70-100% identity, even more preferably 80-100% identity and even more preferably 90-100% identity. In a most preferred embodiment, substantially identical sequences share 95-100% identity. A substantially identical sequence may also relate to a complementary sequence.

Within the sequences of this application, symbols are utilized to identify those

nucleotides that may be represented by more than one of A, T, G, or C. As such, "N" denotes any of A, C, G or T; "R" denotes A or G (purine); "Y" denotes G or T (keto); "M" denotes G or C; and, "W" denotes A or T.

The term *antibody* in its various grammatical forms is used herein to refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antibody combining site or paratope. Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the art as Fab, Fab', $F(ab')_2$ and F(v).

The word *inoculum* in its various grammatical forms is used herein to describe a composition containing a polypeptide of this invention as an active ingredient used for the preparation of antibodies against the polypeptide. When a polypeptide is used in an inoculum to induce antibodies it is to be understood that the polypeptide can be used in various embodiments, e.g., alone or linked to a carrier as a conjugate, or as a polypeptide polymer. However, for ease of expression and in context of a polypeptide inoculum, the various embodiments of the polypeptides of this invention are collectively referred to herein by the term polypeptide and its various grammatical forms.

II. Detection of Nucleic Acids

In one embodiment, the present invention provides for the detection of gene expression where said detected signal is detected as a polynucleotide (such as an RNA, mRNA, DNA, cDNA, or other nucleic acid) or a protein / polypeptide. It should be understood by the skilled artisan that many methods for detection of such signals exist and that any suitable method for detection is encompassed by the instant invention. Typical assay formats utilizing nucleic acid hybridization includes, and are not limited to, 1) nuclear run-on assay, 2) slot blot assay, 3) northern blot assay (Alwine, et al. Proc. Natl. Acad. Sci. 74:5350), 4) magnetic particle separation, 5) Nucleic Acid or DNA chips, 6) reverse northern blot assay, 7) dot blot assay, 8) *in situ* hybridization, 9) RNase protection assay (Melton, et al. Nuc. Acids Res. 12:7035 and as described in the 1998 catalog of Ambion, Inc., Austin, TX), 10) ligase chain reaction, 11) polymerase chain reaction (PCR), 12) reverse transcriptase (RT)-PCR (Berchtold, et al. Nuc. Acids. Res. 17:453), and, 13) differential display RT-PCR (DDRT-PCR) or other suitable

combination of techniques and assays. Methods for detection which can be employed include, and are not limited to 1) radioactive labels, 2) enzyme labels, 3) chemiluminescent labels, 4) fluroescent labels, or other suitable labels. Such methodologies and labels, as well as many other suitable techniques not listed here, are well known in the art and widely available to the skilled artisan.

In an exemplary embodiment, the RNase protection assay may be utilized in the present invention by hybridizing multiple DNA probes corresponding to a one or more members of a panel of sequences to mRNA isolated from a tumor cell and performing the RNase assay. An increase or a decrease in the expression of the sequences from the tumor cell as compared to normal cells indicates that the genes related to those sequences may be involved in tumorigenesis. In a preferred embodiment, the panel is selected from the sequences shown in SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

In another embodiment, multiple DNA probes capable of hybridizing to mRNA corresponding to a reporter sequence under the transcriptional control of a nucleic acid sequence under- or overexpressed in tumor cells transcriptional control region may be utilized. Exemplary reporter sequences may include β-galactosidase, luciferase, CAT, and green fluorescent protein. An increase or a decrease in the expression of the sequences from the tumor cell as compared to normal cells indicates that the genes related to those sequences may be involved in tumorigenesis. In a preferred embodiment, the panel is selected from the sequences of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

The screening assays of the present invention are also well suited for polymerase chain reaction (PCR) amplification, whether the format of such assays are in solution after isolation of mRNA and subsequent direct amplification or such after reverse transcription. Such assays can be performed on isolated biological samples or extracted fluids, using a suitable PCR assay format. The screening methods and compositions of the present invention are also amendable to routine adaptation to automated screening systems employing computer controlled reagent aliquoting and signal detection.

With a known gene target, it is possible to apply standard PCR to assay tissue for specific gene expression (Mok et al., (1994), Gynecologic Oncology, 52: 247-252). However, detection of unknown gene expression requires additional manipulations before a useful gene can be identified. Differential Display Reverse Transcriptase Polymerase Chain Reaction (DDRT-PCR)

is a powerful tool useful for isolating large numbers of expressed nucleic acids, corresponding to gene expression. Several U.S. Patents have been issued relating to methods in this and related methods, including U.S. Patents number 5,599,672; 5,807,680; 5,459,037; 5,814,445; 5,104,792; 4,683,195; 5,665,547; 5,262,311; 5,599,696; and, 5,712,126, to name a few (all of which are hereby incorporated by reference in their entirety). DDRT-PCR has been described by Liang and Pardee (Science, 1993, 257: 967-971); Liang et al. (Nucleic Acids Research, 1993, 21(14): 3269-3275); and, Wang et al. (Trends in Pharmacological Science, 1996, 17(8): 276-9).

Previous attempts to assay brain tumors include the studies of Uchiyama et al. (Neurosurgery, 1995, 37(3): 464-469); Sehgal et al. (J. of Surgical Oncology, 1997, 64: 102-108); Sehgal et al. (Int. J. Cancer, 1997, 71: 565-572); Shinoura et al. (Cancer Letters, 1995, 89: 215-221); and Kito et al. (Gene, 1997, 184: 73-81). However, the direct application of DDRT-PCR to brain tumor samples results in a large number of signals corresponding to expressed genes, not all of which are useful for characterizing the cancerous nature of the brain tumor. Selection of the most significant signals from the large number of signals initially generated, and the assembly of a panel of characteristic nucleic acid targets requires insightful consideration and comparison of the data, followed by re-analysis and assessment of the correctness of such choices. The instant invention provides such a method for the identification of over- or under-expressed sequences in cancer. Preferably, the cancer is of neural origin.

Once identified, the specific nucleic acid targets identified as being characteristic for brain cancer can be readily adapted to automated detection assays for use in diagnosis or screening of patients for predisposition for brain cancer. Modification of the discovery of the unique panel of signals of the present invention for use in such screening or diagnostic assays would be well within the skill of one of ordinary art, and require only routine experimentation.

In one embodiment, detection of a nucleic acid such as an mRNA may be accomplished using a gene chip. For instance, the sequences of interest maybe arrayed upon a chip as described in any of the available gene chip technologies such as that described by Schena, et al. (*Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.* Proc Natl Acad Sci USA, 1996 Oct 1;93(20):10614-9). In that study, DNA "chips" were used to quantitatively monitor differential expression of heat shock and phorbol ester-regulated genes in human T cells. Heller, et al. (*Discovery and analysis of inflammatory disease-related genes using cDNA microarrays.* Proc Natl Acad Sci USA, 1997, Mar 18;94(6):2150-5) used DNA

chips to profile expression of selected human genes of probable significance in inflammation as well as with genes expressed in peripheral human blood cells. In that study, mRNA from cultured macrophages, chondrocyte cell lines, primary chondrocytes, and synoviocytes provided expression profiles for selected cytokines, chemokines, DNA binding proteins, and matrix-degrading metalloproteinases. From the peripheral blood library, tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase genes were identified as expressed differentially in rheumatoid arthritis compared with inflammatory bowel disease. Several other methods for utilizing DNA chips are known, including the methods described in U.S. Patents 5,744,305; 5,733,729; 5,710,000; 5,631,734; 5,599,695; 5,593,839; 5,578,832; 5,556,752; 5,770,722; 5,770,456; 5,753,788; 5,688,648; 5,753,439; 5,744,306 (all of which are incorporated by reference in their entirety).

Adaptation of the teachings of the present invention for nucleic acid or gene chip technology as described above would be routine, following the methods and teachings known in the art. The instant invention provides a DNA chip comprising specific sequences for measuring expression levels of certain sequences within a cancer cell to determine whether expression is up- or down-regulated. For instance, a DNA chip comprising nucleotide sequences capable of hybridizing to one or more members of a panel of DNA sequences may be synthesized using commonly available techniques. mRNA is isolated from a normal, non-cancer cell and a cancer cell and hybridized to the DNA chip comprising one of more of the sequences from the panel. Hybridization is then detected by any of the available methods. In such a manner, sequences that are either overexpressed or underexpressed in a cancer cell as compared to a normal cell are. In a similar manner, mRNA from a cancer cell that has been contacted with a compound may be hybridized to sequences on the DNA chip to determine whether that compound affects expression of a particular sequence. The appropriate controls should be included such that a true comparison can be made. In a preferred embodiment, the members of the panel are selected from the sequences shown in SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

The invention provides for a kit comprising hybridization probes specific for at least two nucleic acid sequences selected from the group consisting of the characteristic nucleic acid sequences that are over- or under-expressed in a cancer cell. Preferably, the sequences are substantially identical to those identified in SEQ ID NOS. 1-184, sequences complementary

thereto, or fragments thereof. In a preferred embodiment, the invention encompasses screening assays for the detection of the expression of at least one of the characteristic nucleic acid sequences identified in SEQ ID NOS. 1-184 below for the diagnosis of potentially cancerous tissues or cells. The invention provides for such a kit, further comprising suitable reaction buffer components. The invention also provides for such a kit wherein said probes are suitable for use in PCR amplification of the specific target, direct or indirect hybridization assay, RNase protection assay. In particular, such screening assays can be performed on tissue biopsy samples, serum samples, cerebro-spinal fluid samples, or any other suitable biological sample.

In another embodiment of the invention, genomic screening assays are contemplated for the detection of specific single nucleotide polymorphisms (SNP) in a nucleic acid sequence found to be over- or under-expressed in a cancer cell. Preferably, the sequence is substantially identical to those listed in SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. In a preferred embodiment, such genomic screening is used to detect any predisposition for cancer formation, as an aid to assist monitoring for potential cancer episodes in the future.

Screening assays for detection of at least one of nucleic acids found to be over- or under-expressed in a cancer cell can be designed on the basis of specific hybridization, under stringent conditions, of at least one probe encompassing a specific nucleic acid sequence. Preferably, the sequence is substantially identical to those of SEQ ID NOS. 1-184, a fragment of such nucleic acid sequence, or as the assay format may require, the complementary nucleic acid sequence, or fragment thereof. The assay can be designed to detect a single species of nucleic acid that is substantially identical to the sequences of SEQ ID NOS. 1-184 in a single assay, or using the properly distinqishable signal mechanisms, more than one specific species per reaction.

In particular, the present invention teaches that the presence of detectable nucleic acid signal corresponding to the nucleic acid sequence of the cDNAs comprising the nucleic acid sequence of one or more of the sequences of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. Thus it is a further aspect of the present invention that the detection of nucleic acid corresponding to novel human genes containing the nucleic acid sequence of one or more of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof as indicative of cancer potential.

III. Methods for Cloning

The identification and isolation of the full-length genes associated with the nucleic acids that found to be over- or under-expressed in a cancer cell provides for the generation of recombinant proteins, via recombinant DNA methodologies, which can be used in numerous ways to prepare and screen for therapeutics that will interact with the protein, such as antibodies and chemical agents. Preferably, the sequence is substantially identical to a sequence of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

A full length polypeptide or fragment thereof encoded by a nucleic acid of the instant invention can be prepared using well known recombinant DNA technology methods such as those set forth in Sambrook et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)) and/or Ausubel et al., eds, (Current Protocols in Molecular Biology, Green Publishers Inc. and Wiley and Sons, N.Y. (1994)). A gene or cDNA encoding protein or fragment thereof may be obtained for example by screening a genomic or cDNA library, or by PCR amplification. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039.

For screening, the probe preferably has a nucleotide sequence corresponding to, complementary to, or substantially identical to a sequence over- or under-expressed in a cancer cell, preferably being a sequences substantially identical to a sequence of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. To probe a cDNA or genomic library using an oligonucleotide probe, the following exemplary hybridization conditions may be utilized: 6X.SSC with 0.05 percent sodium pyrophosphate at a temperature of 35°C-62°C, depending on the length of the oligonucleotide probe. For example, 14 base pair probes may be washed at 35-40°C, 17 base pair probes may be washed at 45-50°C., 20 base pair probes may be washed at 52-57°C, and 23 base pair probes may be washed at 57-63°C. The temperature can be increased 2-3°C where the background non-specific binding appears high. Another exemplary protocol uses tetramethylammonium chloride (TMAC) for the washing step. An exemplary stringent washing solution is 3 M TMAC, 50 mM Tris-HCl, pH 8.0, and 0.2% SDS. As described above, the washing temperature using this solution is a function of the length of the probe (ie, a 17 base pair probe is washed at about 45-50°C).

Alternatively, a gene encoding the polypeptide or fragment may be prepared by chemical synthesis using methods well known to the skilled artisan such as those described by Engels, et

al. (Angew. Chem. Intl. Ed., 28:716-734 (1989)). These methods include, inter alia, the phosphotriester, phosphoramidite, and H-phosphonate methods for nucleic acid synthesis. A preferred method for such chemical synthesis is polymer-supported synthesis using standard phosphoramidite chemistry. Typically, the DNA encoding the polypeptide will be several hundred nucleotides in length. Nucleic acids larger than about 100 nucleotides can be synthesized as several fragments using these methods. The fragments can then be ligated together to form the full length polypeptide. Usually, the DNA fragment encoding the amino terminus of the polypeptide will have an ATG, which encodes a methionine residue. This methionine may or may not be present on the mature form of the polypeptide, depending on whether the polypeptide produced in the host cell is secreted from that cell.

The gene or cDNA so isolated can be inserted into an appropriate expression vector for expression in a host cell. The vector is typically selected to be functional in the particular host cell employed (i.e., the vector is compatible with the host cell machinery such that amplification of the gene and/or expression of the gene can occur). The polypeptide or fragment thereof may be amplified/expressed in prokaryotic, yeast, insect (baculovirus systems) and/or eukaryotic host cells. Selection of the host cell will depend at least in part on whether the polypeptide or fragment thereof is to be glycosylated and/or phosphorylated. If so, yeast, insect, or mammalian host cells are preferable; yeast cells can typically glycosylate and phosphorylate the polypeptide, and insect and mammalian cells can glycosylate and/or phosphorylate the polypeptide as it naturally occurs on the TRIP1 polypeptide (i.e., "native" glycosylation and/or phosphorylation).

Typically, the vectors used in any of the host cells will contain 5' flanking sequence (also referred to as a "promoter") and other regulatory elements as well such as an enhancer(s), an origin of replication element, a transcriptional termination element, a complete intron sequence containing a donor and acceptor splice site, a signal peptide sequence, a ribosome binding site element, a polyadenylation sequence, a polylinker region for inserting the nucleic acid encoding the polypeptide to be expressed, and a selectable marker element.

IV. Methods for Detection of Polypeptides

The invention provides for a method wherein a protein encoded by said expressed gene is detected by protein gel assay, antibody binding assay, or other such detection as is known in the art. For instance, the present invention contemplates a kit comprising specific probes for

detection of a polypeptide product (or fragment thereof) of a sequence that is over- or underexpressed in a cancer cell where such probe can be functionalized antibody protein, polyclonal antibody, monoclonal antibody, or antigen binding fragment of such proteins. Preferably, the nucleic acid encoding the polypeptide or fragment thereof is substantially identical to a sequence of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

An antibody of the present invention, in one embodiment, is characterized as comprising antibody molecules that immunoreact with a protein encoded by a nucleic acid over- or under-expressed in cancer. Preferably, the nucleic acid is substantially identical to a sequence of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. Preferably, an antibody further immunoreacts with the protein in situ, i.e., in a tissue section. Thus, the invention describes an anti-protein antibody that immunoreacts with any of the polypeptides of this invention, preferably also immunoreacts with the recombinant protein corresponding to a nucleic acid of the instant invention, and more preferably also reacts with a native protein in situ in a tissue section.

An antibody of the present invention is typically produced by immunizing a mammal with an inoculum containing a polypeptide of this invention and thereby induce in the mammal antibody molecules having immunospecificity for immunizing polypeptide. The antibody molecules are then collected from the mammal and isolated to the extent desired by well known techniques such as, for example, by using DEAE Sephadex or Protein G to obtain the IgG fraction.

Exemplary antibody molecules for use in the diagnostic methods and systems of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and those portions of an immunoglobulin molecule that contain the paratope, including those portions known in the art as Fab, Fab', $F(ab')_2$ and F(v). Fab and $F(ab')_2$ portions of antibodies are prepared by the proteolytic reaction of papain and pepsin, respectively, on substantially intact antibodies by methods that are well known. See for example, U.S. Patent No. 4,342,566 to Theofilopolous and Dixon. Fab' antibody portions are also well known and are produced from $F(ab')_2$ portions followed by reduction of the disulfide bonds linking the two heavy reduction of the disulfide bonds linking the two heavy chain portions as with mercaptoethanol, and followed by alkylation of the resulting protein mercaptan with a reagent

such as iodoacetamide. An antibody containing intact antibody molecules are preferred, and are utilized as illustrative herein.

The preparation of antibodies against polypeptide is well known in the art. See Staudt et al., J. Exp. Med., 157:687-704 (1983), or the teachings of Sutcliffe, J.G., as described in United States Patent No. 4,900,811, the teaching of which are hereby incorporated by reference. Briefly, to produce a peptide antibody composition of this invention, a laboratory mammal is inoculated with an immunologically effective amount of a polypeptide of this invention typically as present in a vaccine of the present invention. The anti-polypeptide antibody molecules thereby induced are then collected from the mammal and those immunospecific for both a polypeptide and the corresponding recombinant protein are isolated to the extent desired by well known techniques such as, for example, by immunoaffinity chromatography.

To enhance the specificity of the antibody, the antibodies are preferably purified by immunoaffinity chromatography using solid phase-affixed immunizing polypeptide. The antibody is contacted with the solid phase-affixed immunizing polypeptide for a period of time sufficient for the polypeptide to immunoreact with the antibody molecules to form a solid phase-affixed immunocomplex. The bound antibodies are separated from the complex by standard techniques.

For a polypeptide that contains fewer than about 35 amino acid residues, it is preferable to use the peptide bound to a carrier for the purpose of inducing the production of antibodies. One or more additional amino acid residues can be added to the amino- or carboxy-termini of the polypeptide to assist in binding the polypeptide to a carrier. Cysteine residues added at the amino- or carboxy-termini of the polypeptide have been found to be particularly useful for forming conjugates via disulfide bonds. However, other methods well known in the art for preparing conjugates can also be used. The techniques of polypeptide conjugation or coupling through activated functional groups presently known in the art are particularly applicable. See, for example, Aurameas, et al., Scand. J. Immunol., Vol. 8, Suppl. 7:7-23 (1978) and U.S. Patent No. 4,493,795, No. 3,791,932 and No. 3,839,153. In addition, a site-directed coupling reaction can be carried out so that any loss of activity due to polypeptide orientation after coupling can be minimized. See, for example, Rodwell et al., Biotech., 3:889-894 (1985), and U.S. Patent No. 4,671,958. Exemplary additional linking procedures include the use of Michael addition reaction products, di-aldehydes such as glutaraldehyde, Klipstein, et al., J. Infect. Dis., 147:318-326

(1983) and the like, or the use of carbodiimide technology as in the use of a water-soluble carbodiimide to form amide links to the carrier. Alternatively, the heterobifunctional cross-linker SPDP (N-succinimidyl-3-(2-pyridyldithio) proprionate)) can be used to conjugate peptides, in which a carboxy-terminal cysteine has been introduced.

Useful carriers are well known in the art, and are generally proteins themselves. Exemplary of such carriers are keyhole limpet hemocyanin (KLH), edestin, thyroglobulin, albumins such as bovine serum albumin (BSA) or human serum albumin (HSA), red blood cells such as sheep erythrocytes (SRBC), tetanus toxoid, cholera toxoid as well as polyamino acids such as poly D-lysine:D-glutamic acid, and the like. The choice of carrier is more dependent upon the ultimate use of the inoculum and is based upon criteria not particularly involved in the present invention. For example, a carrier that does not generate an untoward reaction in the particular animal to be inoculated should be selected.

A suitable inoculum preferably comprises an effective (i.e., immunogenic) amount of a polypeptide or polypeptide fragment of the present invention, typically as a conjugate linked to a carrier. The effective amount of polypeptide per unit dose sufficient to induce an immune response to the immunizing polypeptide depends, among other things, on the species of animal inoculated, the body weight of the animal and the chosen inoculation regimen is well known in the art. Inocula typically contain polypeptide concentrations of about 10 micrograms (μg) to about 500 milligrams (mg) per inoculation (dose), preferably about 50 micrograms to about 50 milligrams per dose. The term "unit dose" as it pertains to the inocula refers to physically discrete units suitable as unitary dosages for animals, each unit containing a predetermined quantity of active material calculated to produce the desired immunogenic effect in association with the required diluent; i.e., carrier, or vehicle. The specifications for the novel unit dose of an inoculum of this invention are dictated by and are directly dependent on (a) the unique characteristics of the active material and the particular immunologic effect to be achieved, and (b) the limitations inherent in the art of compounding such active material for immunologic use in animals, as disclosed in detail herein, these being features of the present invention.

Inocula are typically prepared from the dried solid polypeptide-conjugate by dispersing the polypeptide-conjugate in a physiologically tolerable (acceptable) diluent such as water, saline or phosphate-buffered saline to form an aqueous composition. Inocula can also include an adjuvant as part of the diluent. Adjuvants such as complete Freund's adjuvant (CFA), incomplete

Freund's adjuvant (IFA) and alum are materials well known in the art, and are available commercially from several sources.

The antibody so produced can be used, inter alia, in the diagnostic methods and systems of the present invention to detect a polypeptide of the present invention in a sample such as a tissue section or body fluid sample. Anti-polypeptide antibodies that inhibit function of the polypeptide can also be used in vivo in therapeutic methods as described herein. A preferred anti-polypeptide antibody is a monoclonal antibody. The phrase "monoclonal antibody" in its various grammatical forms refers to a population of antibody molecules that contain only one species of antibody combining site capable of immunoreacting with a particular epitope. A monoclonal antibody thus typically displays a single binding affinity for any epitope with which it immunoreacts. A monoclonal antibody may therefore contain an antibody molecule having a plurality of antibody combining sites, each immunospecific for a different epitope, e.g., a bispecific monoclonal antibody. A preferred monoclonal antibody of this invention comprises antibody molecules that immunoreact with a polypeptide of the present invention. More preferably, the monoclonal antibody also immunoreacts with recombinantly produced whole protein.

A monoclonal antibody is typically composed of antibodies produced by clones of a single cell called a hybridoma that secretes (produces) only one kind of antibody molecule. The hybridoma cell is formed by fusing an antibody-producing cell and a myeloma or other self-perpetuating cell line. The preparation of such antibodies was first described by Kohler and Milstein, Nature, 256:495-497 (1975), the description of which is incorporated by reference. The hybridoma supernates so prepared can be screened for the presence of antibody molecules that immunoreact with a polypeptide.

Briefly, to form the hybridoma from which the monoclonal antibody composition is produced, a myeloma or other self-perpetuating cell line is fused with lymphocytes obtained from the spleen of a mammal hyperimmunized with a antigen, such as is present in a polypeptide of this invention. The polypeptide-induced hybridoma technology is described by Niman et al., Proc. Natl. Acad. Sci., USA, 80:4949-4953 (1983), the description of which is incorporated herein by reference. It is preferred that the myeloma cell line used to prepare a hybridoma be from the same species as the lymphocytes. Typically, a mouse of the strain 129 GIX^+ is the preferred mammal. Suitable mouse myelomas for use in the present invention include the

hypoxanthine-aminopterin-thymidine-sensitive (HAT) cell lines P3X63-Ag8.653, and Sp2/0-Ag14 that are available from the American Type Culture Collection, Rockville, MD, under the designations CRL 1580 and CRL 1581, respectively. Splenocytes are typically fused with myeloma cells using polyethylene glycol (PEG) 1500. Fused hybrids are selected by their sensitivity to HAT. Hybridomas producing a monoclonal antibody of this invention are identified using the enzyme linked immunosorbent assay (ELISA) described in the Examples.

A monoclonal antibody of the present invention can also be produced by initiating a monoclonal hybridoma culture comprising a nutrient medium containing a hybridoma that produces and secretes antibody molecules of the appropriate polypeptide specificity. The culture is maintained under conditions and for a time period sufficient for the hybridoma to secrete the antibody molecules into the medium. The antibody-containing medium is then collected. The antibody molecules can then be further isolated by well known techniques. Media useful for the preparation of these compositions are both well known in the art and commercially available and include synthetic culture media, inbred mice and the like. An exemplary synthetic medium is Dulbecco's Minimal Essential Medium (DMEM; Dulbecco et al., Virol. 8:396 (1959)) supplemented with 4.5 gm/l glucose, 20 mM glutamine, and 20% fetal calf serum. An exemplary inbred mouse strain is the Balb/c. Other methods of producing a monoclonal antibody, a hybridoma cell, or a hybridoma cell culture are also well known. See, for example, the method of isolating monoclonal antibodies from an immunological repertoire as described by Sastry, et al., Proc. Natl. Acad. Sci. USA, 86:5728-5732 (1989); and Huse et al., Science, 246:1275-1281 (1989).

The monoclonal antibodies of this invention can be used in the same manner as disclosed herein for antibodies of the present invention. For example, the monoclonal antibody can be used in the therapeutic, diagnostic or in vitro methods disclosed herein where immunoreaction with a nucleic acid, polypeptide or fragment thereof, as described herein, is desired. Also contemplated by this invention is the hybridoma cell, and cultures containing a hybridoma cell that produce a monoclonal antibody of this invention.

It is also possible to isolated antibodies reactive against polypeptides of the instant invention using phage display techniques. Display of antibody fragments on the surface of viruses which infect bacteria (bacteriophage or phage) makes it possible to produce human sFvs with a wide range of affinities and kinetic characteristics. To display antibody fragments on the

surface of phage (phage display), an antibody fragment gene is inserted into the gene encoding a phage surface protein (pIII) and the antibody fragment-pIII fusion protein is expressed on the phage surface (McCafferty et al. (1990) Nature, 348: 552-554; Hoogenboom et al. (1991) Nucleic Acids Res., 19: 4133-4137). For example, a sFv gene coding for the V.sub.H and V.sub.L domains of an anti-lysozyme antibody (D1.3) was inserted into the phage gene III resulting in the production of phage with the DI.3 sFv joined to the N-terminus of pIII thereby producing a "fusion" phage capable of binding lysozyme (McCafferty et al (1990) Nature, 348: 552-554). The skilled artisan may also refer to Clackson et al. (1991) Nature, 352: 624-628), (Marks et al. (1992) Bio/Technology, 10: 779-783), Marks et al Bio/Technology, 10: 779-785 (1992) for further guidance. In the instant case, the antibody fragment gene is isolated from the immunized mammal, and inserted into the phage display system. Phage containing antibodies reactive to the polypeptide are then isolated and characterized using well-known techniques. Kits and services are available for generating antibodies by phage display from well-known sources such as Cambridge Antibody Technology Group plc (United Kingdom).

Autoantibodies to the polypeptides of the instant invention may also be detected using techniques well-known and widely available to the skilled artisan. For detection of autoantibodies in the serum of a patient by an antigen-antibody reaction, various conventional immunologically methods can be used such as a method of directly measuring a reaction in a liquid phase and a solid phase and a method of measuring an inhibitory reaction immunologically by adding an inhibiting substance. The following are the examples of the above-mentioned detecting methods, (1) aggregation reaction; (2) DID: double immune diffusion method (Octarony method); (3) ELISA: enzyme linked immunoabsorbent assay, (4) FIA: fluorescent immunosorbent assay, (5) nephlometry method, (6) radioimmuno assay (RIA), (7) immunofluorescent methods. Such methods are described in available references such as U.S. Pat. No. 5,976,810, incorporated herein by reference.

The presence of elevated levels of certain nucleic acids or polypeptides, such as *dek* in gliomas (see below) has potential for development of diagnostic reagents. *dek* has been shown to be an autoantigen in several diseases, such as juvenile rheumatoid arthritis, lupus erythematosis, and Kikuchi's Disease (Szer et al. *A novel autoantibody to the putative oncoprotein DEK in pauciarticular onset juvenile rheumatoid arthritis*. J Rheumatol 1994 Nov;21(11):2136-42; Wichmann et al.. *Autoantibodies to transcriptional regulation proteins*

DEK and ALY in a patient with systemic lupus erythematosus. Hum Immunol 1999 Jan;60(1):57-62; Sierakowska et al. *The putative oncoprotein DEK, part of a chimera protein associated with acute myeloid leukaemia, is an autoantigen in juvenile rheumatoid arthritis*. Clin Exp Immunol 1993 Dec;94(3):435-9; Murray et al. *Antibodies to the 45 kDa DEK nuclear antigen in pauciarticular onset juvenile rheumatoid arthritis and iridocyclitis: selective association with MHC gene*. J Rheumatol 1997 Mar;24(3):560-7; Dong et al. *Autoantibodies to DEK oncoprotein in a patient with systemic lupus erythematosus and sarcoidosis*. Arthritis Rheum 1998 Aug;41(8):1505-10; Arnaudo et al. *Antibodies to the DEK protein in Kikuchi's disease*. J Rheumatol 1998 Sep;25(9):1861-2). The present invention provides for the evaluation of the presence of *dek* autoantibodies in the serum of glioma patients. The existence of such autoantibodies may provide the foundation for both a novel non-invasive diagnostic for gliomas as well as a method for evaluation of tumor recurrence following treatment.

V. Methods of Treatment

a. Pharmacogenomics

The invention further provides for a method of ascertaining propensity for malignancy, monitoring the progress of chemotherapy or other anticancer therapy, screening for re-occurence of cancer, or other similar detection of present or potential cancer, where such method detects for the expression of at least one gene which is over- or under-expressed in a cancer cell. In a preferred embodiment, the gene is nucleic acid sequence sharing substantial identity to a nucleic acid sequence selected from the sequences of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. The present invention provides for a method for ascertaining the propensity for malignant phenotype of cells in a biological sample, said method comprising assaying a biological sample to be tested for a signal indicating the transcription of a nucleic acid transcript, wherein said transcript is from at least one gene selected from the group consisting essentially of the genes encoded for by or containing the characteristic nucleic acid sequences identified in SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

In a further embodiment of the invention, screening assays of biological samples are contemplated, where such assays are conducted during the course of chemotherapy alone, or after surgical intervention to treat cancer, to monitor for the continued presence or return of cancerous cells. Such screening assays are designed to detect for the presence of expressed nucleic acids

corresponding to any of those listed in SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof, as an indicator of the possible tumor recurrence. Such monitoring will quickly identify the effective anti-cancer drugs suitable for treatment of the identified brain cancer. In particular, such methods allow for identifying suitable combination therapies.

Related to the use described above, the methods and compositions of the present invention allow for a therapeutic prediction of the efficacy of any contemplated therapy or therapeutic on the specific brain cancer. By determining the characteristic gene expression features, and testing cells for modulation of such gene expression, it is possible to determine the potential responsiveness of the target brain cancer, to the proposed therapeutic.

Genetic screening is also made possible, as detecting mutations within the genes indicated by the nucleic acid sequences that are over- or underexpressed in a cancer cell. Preferably, the sequences are those in SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. Using the sequences or the control elements of such genes, it is possible to detect and identify persons with a potential predisposition for cancer, and thus bring medical monitoring early in the persons life.

In another embodiment, the present invention provides for a method for monitoring the progression of cancer or the effectiveness of a treatment regimen in a patient. Changes in the expression of certain sequences indicates whether or not a treatment regimen is having an effect in the patient. For example, if a certain treatment regimen results in increased expression of a sequence known to be associated with metastasis, it may be an indication that the treatment is not working to the benefit of the patient.

b. Gene Therapy

The present invention further provides for methods of treating a patient by inhibiting or introducing expression into the cells of a patient a nucleic acid or fragment thereof that shows increased or decreased expression in a tumor cell. The use of gene therapy to augment or ameliorate the expression of the genes associated with the nucleic acid sequences that are over- or under-expressed in tumor cells is also contemplated. In particular, the use of antisense molecules to interfere with mRNAs corresponding to the genes identified by such sequences. It is also possible to construct recombinant DNA vectors which can affect targeted homologous recombination to delete or substitute such genes with normal or non-malignant forms. In a

preferred embodiment, the genes comprise sequence that is substantially identical to the sequences of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

In practicing the present invention, it is advantageous to transfect into a cell a nucleic acid construct directing expression of a protein or nucleic acid product having the ability to alter the behavior of the cell. There are available to one skilled in the art multiple viral and non-viral methods suitable for introduction of a nucleic acid molecule into a target cell. Genetic manipulation of primary tumor cells has been described previously (Patel et al., 1994. Human Gene Therapy 5, p. 577-584). Genetic modification of a cell may be accomplished using one or more techniques well known in the gene therapy field (Human Gene Therapy April 1994, Vol. 5, p. 543-563; Mulligan, R.C. 1993). Viral transduction methods may comprise the use of a recombinant DNA or an RNA virus comprising a nucleic acid sequence that drives or inhibits expression of a protein to infect a target cell. A suitable DNA virus for use in the present invention includes but is not limited to an adenovirus (Ad), adeno-associated virus (AAV), herpes virus, vaccinia virus or a polio virus. A suitable RNA virus for use in the present invention includes but is not limited to a retrovirus or Sindbis virus. It is to be understood by those skilled in the art that several such DNA and RNA viruses exist that may be suitable for use in the present invention.

Adenoviral vectors have proven especially useful for gene transfer into eukaryotic cells (Stratford-Perricaudet, L., and M. Perricaudet. 1991. *Gene transfer into animals: the promise of adenovirus.* p. 51-61, In: Human Gene Transfer, Eds, O. Cohen-Haguenauer and M. Boiron, Editions John Libbey Eurotext, France). Adenoviral vectors have been successfully utilized to study eukaryotic gene expression (Levrero, M., et al. 1991. *Defective and nondefective adenovirus vectors for expressing foreign genes in vitro and in vivo.* Gene 101: 195-202), vaccine development (Graham, F. L., and L. Prevec (1992) *Adenovirus-based expression vectors and recombinant vaccines. In* Vaccines: New Approaches to Immunological Problems, (Ellis, R. V. Ed.), pp. 363-390. Butterworth-heinemann, Boston), and in animal models (Stratford-Perricaudet, et al. 1992. *Widespread long-term gene transfer to mouse skeletal muscles and heart.* J. Clin. Invest. 90, 626-630; Rich, et al. 1993. *Development and analysis of recombinant adenoviruses for gene therapy of cystic fibrosis.* Human Gene Ther. 4, 461-476). The first trial of Ad-mediated gene therapy in human was the transfer of the cystic fibrosis transmembrane conductance regulator (CFTR) gene to lung (Crystal, et al. 1994. Nature Genetics 8, 42-51).

Experimental routes for administrating recombinant Ad to different tissues *in vivo* have included intratracheal instillation (Rosenfeld, et al. 1992. *In vivo transfer of the human cystic fibrosis transmembrane conductance regulator gene to the airway epithelium.* Cell 68, 143-155) injection into muscle (Quantin, B., et al. 1992. *Adenovirus as an expression vector in muscle cells in vivo.* Proc. Natl. Acad. Sci. USA 89, 2581-2584), peripheral intravenous injection (Herz, J. and R.D. Gerard. 1993. *Adenovirus-mediated transfer of low density lipoprotein receptor gene acutely accelerates cholesterol clearance in normal mice.* Proc. Natl. Acad. Sci. USA 90, 2812-2816) and stereotactic inoculation to brain (Le Gal La Salle, et al. 1993. *An adenovirus vector for gene transfer into neurons and glia in the brain.* Science 259, 988-990). The adenoviral vector, then, is widely available to one skilled in the art and is suitable for use in the present invention.

Adeno-associated virus (AAV) has recently been introduced as a gene transfer system with potential applications in gene therapy.. Wild-type AAV demonstrates high-level infectivity, broad host range and specificity in integrating into the host cell genome (Hermonat, P.L., and N. Muzyczka. 1984. *Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells.* Proc. Natl. Acad. Sci. USA 81: 6466-6470). Herpes simplex virus type-1 (HSV-1) is attractive as a vector system for use in the nervous system because of its neurotropic property (Geller, A.I., and H.J. Federoff. 1991. *The use of HSV-1 vectors to introduce heterologous genes into neurons: implications for gene therapy.* In: Human Gene Transfer, Eds, O. Cohen-Haguenauer and M. Boiron, pp. 63-73, Editions John Libbey Eurotext, France; Glorioso, et al. 1995. *Herpes simplex virus as a gene-delivey vectors for the central nervous system.* In: Viral Vectors-Gene therapy and neuroscience application, Eds, M.G. Kaplitt and A.D. Loewy, pp. 1-23. Academic Press, New York). Vaccinia virus, of the poxvirus family, has also been developed as an expression vector (Smith, G.L., and B. Moss. 1983. *Infectious poxvirus vectors have capacity for at least 25,000 base pairs of foreign DNA.* Gene 25: 21-28; Moss, B. 1992. *Poxviruses as eukaryotic expression vectors.* Semin. Virol. 3: 277-283; Moss, B. 1992. *Poxviruses as eukaryotic expression vectors.* Semin. Virol. 3: 277-283). Each of the above-described vectors are widely available to one skilled in the art and would be suitable for use in the present invention.

Retroviral vectors are capable of infecting a large percentage of the target cells and integrating into the cell genome (Miller, A.D., and G.J. Rosman. 1989. *Improved retroviral vectors for gene therapy and expression.* Biotechniques 7: 980-990). Retroviruses were

developed as gene transfer vectors relatively earlier than other viruses, and were first used successfully for gene marking and transducing the cDNA of adenosine deaminase (ADA) into human lymphocytes.

It is also possible to produce a viral vector in vivo by implantation of a "producer cell line" in proximity to the target cell population. As demonstrated by Oldfield, et al. (*Gene Therapy for the Treatment of Brain Tumors Using Intra-Tumoral Transduction with the Thymidine Kinase Gene and Intravenous Ganciclovir*, Human Gene Therapy 4:39-69), infiltration of a brain tumor with cells engineered to produce a viral vector carrying an effector gene results in the continuous release of the viral vector in the vacinity of the tumor cells for an extended period of time (i.e, several days). In such a system, the vector is retroviral vector which preferably infects proliferating cells, which, in the brian, would include mainly tumor cells. The present invention provides a methodology with which a viral vector supplies a nucleic acid sequence encoding a protein having sialyltransferase activity to cells involved in a nuerological disorder such as brain cancer.

"Non-viral" delivery techniques that have been used or proposed for gene therapy include DNA-ligand complexes, adenovirus-ligand-DNA complexes, direct injection of DNA, $CaPO_4$ precipitation, gene gun techniques, electroporation, and lipofection (Mulligan, R.C. 1993. *The basic science of gene therapy.* Science 260: 926-932). Any of these methods are widely available to one skilled in the art and would be suitable for use in the present invention. Other suitable methods are available to one skilled in the art, and it is to be understood that the present invention may be accomplished using any of the available methods of transfection. Several such methodologies have been utilized by those skilled in the art with varying success (Mulligan, R.C. 1993. *The basic science of gene therapy.* Science 260: 926-932). Lipofection may be accomplished by encapsulating an isolated DNA molecule within a liposomal particle and contacting the liposomal particle with the cell membrane of the target cell. Liposomes are self-assembling, colloidal particles in which a lipid bilayer, composed of amphiphilic molecules such as phosphatidyl serine or phosphatidyl choline, encapsulates a portion of the surrounding media such that the lipid bilayer surrounds a hydrophilic interior. Unilammellar or multilammellar liposomes can be constructed such that the interior contains a desired chemical, drug, or, as in the instant invention, an isolated DNA molecule.

The cells may be transfected in vivo (preferably at the tumor site), ex vivo (following

removal from a primary or metastatic tumor site), or in vitro. The cells may be transfected as primary cells isolated from a patient or a cell line derived from primary cells, and are not necessarily autologous to the patient to whom the cells are ultimately administered. Following ex vivo or in vitro transfection, the cells may be implanted into a host, preferably a patient having a neurological disorder and even more preferably a patient having a brain tumor. Genetic manipulation of primary tumor cells has been described previously (Patel et al., 1994. Human Gene Therapy 5, p. 577-584). Genetic modification of the cells may be accomplished using one or more techniques well known in the gene therapy field (Human Gene Therapy. April 1994. Vol. 5, p. 543-563; Mulligan, R.C. 1993. *The basic science of gene therapy.* Science 260: 926-932).

In order to obtain transcription of the nucleic acid of the present invention within a target cell, a transcriptional regulatory region capable of driving gene expression in the target cell is utilized. The transcriptional regulatory region may comprise a promoter, enhancer, silencer or repressor element and is functionally associated with a nucleic acid of the present invention. Preferably, the transcriptional regulatory region drives high level gene expression in the target cell. It is further preferred that the transcriptional regulatory region drives transcription in a cell involved in a neurological disorder such as brain cancer. Transcriptional regulatory regions suitable for use in the present invention include but are not limited to the human cytomegalovirus (CMV) immediate-early enhancer/promoter, the SV40 early enhancer/promoter, the JC polyomavirus promoter and the chicken β-actin promoter coupled to the CMV enhancer (Doll,et al. 1996. *Comparison of promoter strengths on gene delivery into mammalian brain cells using AAV vectors.* Gene Therapy 3: 437-447). Other transcriptional regulatory regions useful for practicing the present invention are available and well known in the art, and are contemplated as being part of the present invention.

The vectors of the present invention may be constructed using standard recombinant techniques widely available to one skilled in the art. Such techniques may be found in common molecular biology references such as *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, CA), and *PCR Protocols: A Guide to Methods and Applications* (Innis, et al. 1990. Academic Press, San Diego, CA). Examples of nucleic acid constructs useful for practicing the present invention comprise

a transcriptional regulatory region such as the CMV immediate-early enhancer/promoter, the SV40 early enhancer/promoter, the JC polyomavirus promoter, or the chicken β-actin promoter coupled to the CMV enhancer operably linked to a nucleic acid comprising one or more of SEQ ID NOS. 1-184, or a fragment or complement thereof. To generate such a construct, a nucleic acid sequence encoding the enzyme may be processed using one or more restriction enzymes such that certain sequences flank the nucleic acid. Processing of the nucleic acid may include the addition of linker or adapter sequences. A nucleic acid sequence comprising a preferred transcriptional regulatory region may be similarly processed such that the sequence has flanking sequences compatible with the nucleic acid sequence encoding the enzyme. These nucleic acid sequences may then be joined into a single construct by processing of the fragments with an enzyme such as DNA ligase. The joined fragment, comprising a transcriptional regulatory region operably linked to a nucleic acid comprising a sequence that is over- or underexpressed in a cancer cell, preferably being a sequence substantially identical to a sequence of SEQ ID NOS. 1-184, or a fragment or complement thereof, may then be inserted into a plasmid capable of being replicated in a host cell by further processing using one or more restriction enzymes.

Administration of a nucleic acid of the present invention to a target cell in vivo may be accomplished using any of a variety of techniques well known to those skilled in the art. Such reagents may be administered by intravenous injection or using a technique such as stereotactic injection to administer the reagent into the target cell or the surrounding areas (Badie, et al. 1994. *Stereotactic Delivery of a Recombinant Adenovirus into a C6 Glioma Cell Line in a Rat Brain Tumor Model.* Neurosurgery 35: 910; Perez-Cruet, et al. 1994. *Adenovirus-Mediated Gene Therapy of Experimental Gliomas.* J. Neur. Res. 39: 506; Chen, et al. 1994. *Gene therapy for brain tumors: Regression of experimental gliomas by adenovirus-mediated gene transfer in vivo.* Proc. Natl. Acad. Sci. USA 91: 3054; Oldfield, et al. 1993. *Gene Therapy for Treatment of Brain Tumors Using Intra-Tumoral Transduction with the Thymidine Kinase Gene and Intravenous Ganciclovir.* Human Gene Therapy 4:39-69; Okada, et al. 1996).

In another embodiment, the present invention provides a methodology for transfection of a functional nucleic acid sequence, preferably an antisense oligonucleotide, that inhibits expression of a nucleic acid comprising a sequence of SEQ ID NOS. 1-184, or a protein encoded by a nucleic acid comprising a sequence of SEQ ID NOS. 1-184. The antisense oligonucleotide may comprise a functional nucleotide sequence such as a 2',5'-oligoadenylate as described in

U.S. Patent No. 5,583,032. Using such an antisense oligonucleotide, expression of a protein comprising a sequence substantially identical to that encoded by the sequences of SEQ ID NOS. 1-184 may be inhibited by inhibition of transcription, destruction of the transcript encoding the protein, or inhibition of translation of the protein from its transcript.

In certain embodiments of the present invention, transfection of a cell is performed. In a preferred embodiment, the cell is involved in the causation of a neurological disorder such as brain cancer, Parkinson's disease or Alzheimer's disease. In a preferred embodiment, the cell is a cancer cell, and in a more preferred embodiment, the cell is a brain cancer cell. More preferably, the nucleic acid comprises a sequence encoding the protein encoded by a nucleic acid comprising a sequence of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof is under the transcriptional control of a transcriptional regulatory region which functions within a neural tissue or cell.

In another embodiment of the present invention, a target cell is transfected in vivo by implantation of a "producer cell line" in proximity to the target cell population (Oldfield, et al. 1993. *Gene Therapy for Treatment of Brain Tumors Using Intra-Tumoral Transduction with the Thymidine Kinase Gene and Intravenous Ganciclovir.* Human Gene Therapy 4:39-69; Culver, et al. 1994. *Gene Therapy for the Treatment of Malignant Brain Tumors with in vivo Tumor Transduction with the Herpes Simplex Thymidine Kinase Gene/Ganciclovir System,* Human Gene Therapy 5: 343-379). The producer cell line is engineered to produce a viral vector and releases viral particles in the vicinity of the target cell. A portion of the released viral particles contact the target cells and infect those cells, thus delivering a nucleic acid of the present invention to the target cell. Following infection of the target cell, expression of the product of nucleic acid of the present invention occurs. Preferably, expression results in either increased or decreased expression of a protein encoded by the nucleic acid, which preferably comprises substantially identical DNA sequence to the sequences of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

In yet another embodiment, the present invention comprises a kit for determining the tumorigenicity or malignancy of a brain cell. The kit may comprise a panel of independent or paired nucleic acid molecules specific for the detection of the expression of specific nucleic acid sequences corresponding to nucleic acid sequences that are over- or under-expressed in cancer cells. Preferably, the sequences are substantially identical to those of SEQ ID NOS. 1-184,

sequences complementary thereto, or fragments thereof. One embodiment of such a kit utilizes enzyme-mediated nucleic acid amplification such as the polymerase chain reaction (PCR) in which a pair of nucleic acid molecules (i.e., primers) that allow for amplification of a nucleic acid sequence of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof.

c. Small Molecules

The methods and compositions of the present invention are useful for the manufacture of pharmaceuticals and therapeutics which encompass compounds that interact with or affect the expression of nucleic acid sequences or proteins over- or underexpressed in cancer cells. Preferably, the nucleic acid sequences comprise sequence substantially identical to those sequences listed in SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. Such inhibitors can take the form of traditional chemotherapeutic agents, as well as specific anti-sense nucleic acids targeted to the nuclei acid sequences. Such therapeutics can be directed against single nucleic acid targets, but most preferably are targeted at more than one specific nucleic acid sequence.

The present invention also provides for therapeutic compounds identified or otherwise identifiable by this method, and any compound corresponding to a compound identified by these methods. The reagents and methodologies of the present invention provide an assay system for determining the effect of a compound on gene expression in a cell. In one embodiment, the cell may be affected such that upon administration of the compound to a patient, cell growth or activity that may be detrimental to the patient may result. In such cases, it would be beneficial to have at the researcher's disposal a rapid, accurate, and efficient assay system to measure the likelihood that a compound may have such effects. Preferably, the "panel" refers to the sequences substantially identical to one or more of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. It is to be understood by the skilled artisan that the present invention provides an assay or test system that is applicable to many types of cells and panels of nucleotide sequences.

In one embodiment, the present invention provides an assay for identifying a compound that may promote or prevent cancer. A method for identifying a compound affecting a cell is provided wherein a cell is contacted with a compound and expression of one or more nucleotide sequences or proteins selected from a panel of sequences is detected. The panel may consist of

one or more sequences of the invention. The level of expression may be compared to control levels, such as where a cell has not been contacted by the compound but is otherwise maintained under identical conditions as the cell that has been contacted. In one embodiment, a method for detecting a compound that may promote cancer comprising detection of increased expression of the panel of sequences following contact of the cell with the compound is provided. In another embodiment, a method for detecting decreased expression of one or more members of the panel of sequences following exposure to the compound, thus identifying a compound that may inhibit tumor cell migration. In yet another embodiment, a method for detecting increased expression of the one or more members of the panel following exposure to the compound, thus identifying a compound that may promote tumor cell migration. In a preferred embodiment, the present invention provides an assay for identifying a compound that may promote or prevent brain cancer. In one embodiment, the sequences are selected from sequences substantially identical to those sequences in SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. Any combination of such sequences may be combined to provide a useful assay system as described herein.

In one embodiment of the present invention, a method for identifying a compound affecting a cell is provided wherein a cell is contacted with a compound and expression of a reporter gene functionally linked to a transcriptional regulatory sequence of a nucleotide sequence that is up- or down-regulated in cancer cells. In a preferred embodiment, the reporter sequences is β-galactosidase, luciferase, green fluorescent protein or chloramphenicol acetyl transferase (CAT). In a preferred embodiment, the transcriptional regulatory region controls the expression of a sequence substantially identical to a sequence of SEQ ID NOS. 1-184, sequences complementary thereto, or fragment thereof.

In yet another embodiment, the present invention comprises a kit for determining the effect of a compound on gene expression within a cell. The kit may comprise packaged reagents such as a panel of independent or paired nucleic acid molecules specific for the detection of the expression of specific nucleic acid sequences corresponding to specific species of nucleic acid sequences encoding proteins of interest. Instructions for use of the packaged reagent(s) are also typically included. As used herein, the term "package" refers to a solid matrix or material such as glass, plastic (e.g., polyethylene, polypropylene or polycarbonate), paper, foil and the like capable of holding within fixed limits a polyamide of the present invention. "Instructions for use"

typically include a tangible expression describing the reagent concentration or at least one assay method parameter such as the relative amounts of reagent and sample to be admixed, maintenance time periods for reagent or sample admixtures, temperature, buffer conditions and the like.

In another embodiment, the present invention provides a compound identified by its ability to cause an increase or a decrease in one or more sequences of a panel of sequences. The compounds of this invention may be formulated into diagnostic and therapeutic compositions for *in vivo* or *in vitro* use. Representative methods of formulation may be found in *Remington: The Science and Practice of Pharmacy*, 19th ed., Mack Publishing Co., Easton, PA (1995). For *in vivo* use, the compound may be incorporated into a physiologically acceptable pharmaceutical composition that is administered to a patient in need of treatment or an animal for medical or research purposes. The polyamide composition comprises pharmaceutically acceptable carriers, excipients, adjuvants, stabilizers, and vehicles. The composition may be in solid, liquid, gel, or aerosol form. The polyamide composition of the present invention may be administered in various dosage forms orally, parentally, by inhalation spray, rectally, or topically. The term parenteral as used herein includes, subcutaneous, intravenous, intramuscular, intrasternal, infusion techniques or intraperitoneally.

The selection of the precise concentration, composition, and delivery regimen is influenced by, *inter alia*, the specific pharmacological properties of the particular selected compound, the intended use, the nature and severity of the condition being treated or diagnosed, the age, weight, gender, physical condition and mental acuity of the intended recipient as well as the route of administration. Such considerations are within the purview of the skilled artisan. Thus, the dosage regimen may vary widely, but can be determined routinely using standard methods.

The pharmaceutically active compounds (i.e., polypeptides, nucleic acids, compounds or vectors) of this invention can be processed in accordance with conventional methods of pharmacy to produce medicinal agents for administration to patients, including humans and other mammals. For oral administration, the pharmaceutical composition may be in the form of, for example, a capsule, a tablet, a suspension, or liquid. The pharmaceutical composition is preferably made in the form of a dosage unit containing a given amount of DNA or viral vector particles (collectively referred to as "vector"). For example, these may contain an amount of

vector from about 10^3-10^{15} viral particles, preferably from about 10^6-10^{12} viral particles. A suitable daily dose for a human or other mammal may vary widely depending on the condition of the patient and other factors, but, once again, can be determined using routine methods. The vector may also be administered by injection as a composition with suitable carriers including saline, dextrose, or water.

Injectable preparations, such as sterile injectable aqueous or oleaginous suspensions, may be formulated according to the known are using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation may also be a sterile injectable solution or suspension in a non-toxic parenterally acceptable diluent or solvent, for example as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil may be employed, including synthetic mono- or diglycerides. In addition, fatty acids such as oleic acid find use in the preparation of injectables.

A suitable topical dose of active ingredient of a vector of the present invention is administered one to four, preferably two or three times daily. For topical administration, the vector may comprise from 0.001% to 10% w/w, *e.g.*, from 1% to 2% by weight of the formulation, although it may comprise as much as 10% w/w, but preferably not more than 5% w/w, and more preferably from 0.1% to 1% of the formulation. Formulations suitable for topical administration include liquid or semi-liquid preparations suitable for penetration through the skin (*e.g.*, liniments, lotions, ointments, creams, or pastes) and drops suitable for administration to the eye, ear, or nose.

The pharmaceutical compositions may be made up in a solid form (including granules, powders or suppositories) or in a liquid form (*e.g.*, solutions, suspensions, or emulsions). The pharmaceutical compositions may be subjected to conventional pharmaceutical operations such as sterilization and/or may contain conventional adjuvants, such as preservatives, stabilizers, wetting agents, emulsifiers, buffers etc. Solid dosage forms for oral administration may include capsules, tablets, pills, powders, and granules. In such solid dosage forms, the active compound may be admixed with at least one inert diluent such as sucrose, lactose, or starch. Such dosage forms may also comprise, as in normal practice, additional substances other than inert diluents, *e.g.*, lubricating agents such as magnesium stearate. In the case of capsules, tablets, and pills,

the dosage forms may also comprise buffering agents. Tablets and pills can additionally be prepared with enteric coatings. Liquid dosage forms for oral administration may include pharmaceutically acceptable emulsions, solutions, suspensions, syrups, and elixirs containing inert diluents commonly used in the art, such as water. Such compositions may also comprise adjuvants, such as wetting sweetening, flavoring, and perfuming agents.

The compositions of the present invention may be administered orally, parentally, by inhalation spray, rectally, or topically in dosage unit formulations containing conventional pharmaceutically acceptable carriers, adjuvants, and vehicles. The term parenteral as used herein includes, subcutaneous, intravenous, intramuscular, intrasternal, infusion techniques or intraperitoneally. Suppositories for rectal administration of the drug can be prepared by mixing the drug with a suitable non-irritating excipient such as cocoa butter and polyethylene glycols that are solid at ordinary temperatures but liquid at the rectal temperature and will therefore melt in the rectum and release the drug.

The dosage regimen for compositions of this invention is based on a variety of factors, including the type of disease, the age, weight, sex, medical condition of the patient, the severity of the condition, the route of administration, and the particular compound employed. Thus, the dosage regimen may vary widely, but can be determined routinely using standard methods.

While the compounds, polypeptides, nucleic acids and /or vectors of the invention can be administered as the sole active pharmaceutical agent, they can also be used in combination with one or more vectors of the invention or other agents. When administered as a combination, the therapeutic agents can be formulated as separate compositions that are given at the same time or different times, or the therapeutic agents can be given as a single composition.

VI. Conclusions

Thus the compositions and methods of the present invention are useful as clinical screens for the specific diagnosis and identification of cancer. Preferably, the cancer is brain cancer, and more preferably, the cancer is glioma. In one embodiment, the strong indication of glioma is characterized by detection of increased or decreased expression of SEQ ID NOS. 1-184, sequences complementary thereto, or fragments thereof. The methods and assays of the invention are also useful for the detection of potential cancer development such as glioma or other cancers. Thus the determination and early detection of glioma propensity greatly assists

the medical practitioner and patient decide upon the proper course of action. Once such action is taken, the methods of the present invention allows for the monitoring of recurrence after surgery, or during the course of chemotherapy.

The following Examples are for illustrative purposes only and are not intended, nor should they be construed as limiting the invention in any manner. Those skilled in the art will appreciate that variations and modifications can be made without violating the spirit or scope of the invention.

EXAMPLES

As discussed above, DDRT-PCR is a powerful method for identifying and analyzing altered gene expression at the mRNA level. It has been utilized to identify cellular mRNAs whose expression is altered in malignant brain tumors, and has successfully yielded several genes. Most of these to date are still of unknown function and clinical utility. Established herein is a reliable DDRT-PCR/screening protocol to study modulation of gene expression in human brain tumors. A comparison between cultured NHFA and a tumorigenic glioma cell line, U373MG was initially chosen for study. This system provided a proliferative model of glial lineage which supplied both well-defined and renewable resources necessary for our intensive screening protocols. Following DDRT-PCR using a panel of 84 unique primer pairs, differentially expressed amplicons were further screened by a series of Northern analyses. As described below, comparison of cultured normal human fetal astrocytes (NHFA) with a tumorigenic glioma cell line (U373MG) initially generated at least 142 differentially expressed transcripts, wherein at least SEQ ID NOS. 1-94 appeared to be under-expressed in the tumor cells. In addition, at least SEQ ID NOS. 95-141 and 183 appear to be over-expressed in tumor cells. SEQ ID NO. 68, 69 and 183 were further confirmed by reverse northern blot.

Age at primary diagnosis is among the most significant factors impacting survival of patients with glioblastomas (GBM). Patients diagnosed prior to the age of 50 years survive significantly longer than those diagnosed after the age of 50, with median survival of 24 months and 8 months, respectively. This differential survival is independent of performance status and appears to be unrelated to treatment. The cellular mechanisms for this age/prognosis correlation are not known. Several age-related genetic alterations have been recently demonstrated in malignant gliomas, suggesting that there is a molecular basis for this poor patient survival.

Overall survival of patients diagnosed with GBMs demonstrates a marked inverse age-dependence (**Figure 1**). In order to understand the molecular basis for this poor patient survival, we utilized a DDRT-PCR-based strategy and identified multiple differentially expressed mRNAs in GBMs excised from older (>60 yr.) and younger (<45 yr.) patients As shown below, DDRT-PCR indicates that SEQ ID NOS. 142-174 are over-expressed in tumors from old patients as compared to those of young patients. SEQ ID NOS: 175-182 were determined to be under-expressed in tumors of old patients as compared to those of young patients. The expression of SEQ ID NOS. 142, 143, 144, 147, 149, 162 and 173 were confirmed by reverse northern blot.

Example 1

Isolation of RNA

Human glioblastoma cell line U373MG (American Type Culture Collection - ATCC, Manassas, VA) was the source of malignant phenotype expression signals. Cultured normal human fetal astrocytes, isolated according to Yamamoto et al., (1997, Brain Research 755(1):175-9), and processed no later than 20 passages from the initial isolation, was the source of normal tissue expression signals. All cells were subcultured in Dulbeccos Modified Eagles Medium (DMEM) containing 10% heat-inactivated fetal bovine serum (FBS; Whittaker BioProducts, Walkersville, MD), penicillin/streptomycin and glutamine and were maintained in log phase at 37°C in the presence of 10% CO_2.

The material for the secondary clinical reverse northern screens was obtained with informed consent from two sources: (1) normal human brain tissue was obtained from the Brain and Tissue Bank for Developmental Disorders at the University of Maryland, (Baltimore, MD), (2) human brain tumor tissue, from donor tissue, glioblastoma multiforme, recurrent glioblastoma multiforme, and astrocytoma grade IV (glioblastoma) was obtained from excised tumor material. The clinical material,as classified according to WHO Brain Tumor Classification, are all treated as glioblastoma tissue.

Briefly, total RNA was extracted from tissues by guanidinium thiocyanate treatment, followed by separation using cesium chloride centrifugal sedimentation, and treated with DNase I for 30 minutes at 37°C.

RT-PCR was performed on the extracted RNA using commercially availible oligonucleotide primers, following the recommended procedures. Specifically, anchored primers

and 20 arbitrary 10-mer primers from Operon Technologies, Inc. (Kit A; Alemeda, CA), and 8 arbitrary 13-mer primers from GenHunter Corp. (Cat. No. H-AP-D; Brookline, MA) were selected. Specifically, the primers were:

Anchored Primer $T_{11}M$ (where M is A, C or G)

Random Primer Operon Technologies, Kit A, primers OPA-01 to OPA-20

GenHunter Corp., H-AP primer set 4, primers H-AP25 to H=AP32

The combination of the primers from these two commercial kits produce a total of 84 unique primer pairs. Differential display was performed essentially as described by Liang et al., (Science, 1992, 257:967-71). For each of the three anchored primers in each sample, 0.2 ug of total DNA-free RNA was reverse transcribed with 50U Maloney Murine Lukemia Virus (MMLV) reverse transcriptase in the presence of 200 pmol anchored primer, and 20 uM dNTP for 5 minutes at 65°C, followed by 60 minutes at 37°C. Following heat inactivation of the reverse transcriptase at 75°C for 5 minutes, 2 µl of the RT mixture was amplified in the presence of 2 uM dNTP, 200 nM of the appropriate anchored primer, 4 pmol arbitrary (random) primer, 10 uCi α-[^{32}P]dATP (1000-3000 Ci/mmol; Amersham Corp., Arlington Hts., IL), and 1 Unit of AmpliTaq$^{(R)}$ (*T. aqut.* DNA polymerase; Perkin-Elmer Corp., Branchburg, NJ). The cycling parameters were: 94°C for 15 sec., 40°C for 2 min., 72°C for 30 sec., for 40 cycles. Following a final extension for 5 min. at 72°C, the samples were stored at 4°C until analysis. The PCR products were electrophoresed on 6% sequencing gels. Differentially expressed bands of interest were excised from the dried gel, boiled in dH_2O, purified by ethanol precipitation, and reconstituted in 10 µl dH_2O.

The minimal selection criteria for the bands of interest was approximately two-fold greater signal expressed in either tissue, and was qualitatively evaluated by visual inspection of the autoradiographic image.

Example 2

Characterization of Sequences

An aliquot of the purified cDNA amplicons were then reamplified and subcloned into the cloning site of a cloning vector and insert-containing vectors from multiple positive transformants were sequenced using an ABI 377 automated fluorescence-based nucleic acid sequencer. All NCBI maintained nucleotide databases (National Center for Biotechnology

Information; Bethesda, MD) were searched for homologies using the BLAST (basic local alignment search tool) program. The following sequences were identified as being under-expressed in tumor tissue as compared to normal tissue (SEQ ID NOS. 1-94) or over-expressed in tumor tissues as compared to normal tissues (SEQ ID NOS. 95-141).

SEQ ID NO. 1: **NA1-1-N**

TCAGGCCCTTCATGTTAGTAAAAGCAGACAGACTTTTATATAAAGCCCAGCTTTACCTTTTACTTATTAGTTTGA
ATGAACTTGGGCAAGTTACTTAGTTTTCTGAATCTCATTTTTTCAAATGAAAATTAATTCCATATAATTCCTTCT
CTAGGGGATTTAATTATTATTAGAGACAGGGTCTCACTGTGTCACATAGGCTGGAGTGCAGTGGTGTGATCATAG
CTCATCGTATCCTCAAACTACTTGGCTCAAGCAATCCCCCTGCCTCAGCCTCCTAAGTAGCTAGGACTACAGGCG
TGTGCCACCTTGCCTGGCTAACTAAAAAAAAAAAGCTT

SEQ ID NO. 2: **NA1-1-P**

CAGGCCCTTCCAAAAAAATAGAAGTGGAGGAAACAATTCCTAACACATTCCTTGAGGCCAGCATTACCGTGGTAG
CTGAGCCCGATAAAAATGGTCATAGAAGAGAAAATCACAAACCATATCCCTTATCAATGTAGATGCTAAAATTTT
CCACAGAATACCAGCAAACTTAATCCAACAGTGTATTAAAAGGTTTAGACTTGTCATCAGGTGGGATTTATTCCA
GGAATGTAAAAGTGGTTCAGTTTAAGAAAATTAATTAACACTACCTGCACATCTCAGTTGACACACGAAAGGTGT
CTGACAAAATCTCATAACTGTTCATGATAAAAAAAAAAAGCTT

SEQ ID NO. 3: **NA2-1-F,G,H**

TTGCCGAGCTGGAATTGGAAAGAAGGTGATGACGCAATCTGCCTCGCAGAGTTGAAGTTGGGCTTCATAGCCCAG
AGCTGCCTGGCTCAAGGCCTCTCCACCATGCTTGCCAACCTCTTCTCCATGAGGTCATTCATAAAGATTGAGGAA
GACACATGGCAGAAATACTACTTGGAAGGAGTCTCAAATGAAATGTACACAGAATATCTCTCCAGTGCCTTCGTG
GGTCTGTCCTTCCCTACTGTTTGTGAGCTGTGTTTTGTGAAGCTCAAGCTCCTAATGATAGCCATTGAGTACAAG
TCTGCCAACCGAGAGAGCCGAAGCCGAAAGCGTATATTAATTAATCCTGGAAACCATCTTAAGATCCAAGAAGGT
ACTTTAGGATTTTTCATCGCAAGTGATGCCAAAGAAGTTAAAAGGGCATCTTTTTACTGCAAGGCCTGTCATGAT
GACATCACAGATCCCAAAAGAATAAAAAAATGTGGCTGCAAACGGCTTGAAGATGAGCAGCCCGTCAACACTATC
ACCAAAAAAAAAAAAGCTTT

SEQ ID NO. 4: **NA5-1-F,H**

AGGGGTCTTGCAGAATGGAATTAACCTGAATTCAACAAAAGAGGTCTTTAAAATTCATAACAGCAGGTGTCGTCT
GTCTTTGAGATTCCCTTGCCAAAAAAGGAAATGATTTCTTAGTGATATGCTTTACTTCTGTTGATCACTATTTGC
TCTTTTAAAGTGTCCAAAGATGTTTTAATAGATACTTGGTATTTGTTGTTTTCTTTAATAAAGTATAATTTACAT
GTAAAAAAAAAAAGCTT

SEQ ID NO. 5: **NA5-1-G**

AGGGGTCTTGGCACAGGAAAAGGACAGTAGGTCAAAACTAAGGAATATCAATGAAGTATGGGCCTTAGTTAATAT
TAAAGTATCAATATTGGTATATTAGTTGTATCAAATGTATCATACTAATGTAAGATATTAACCATAGGGAGAACT
GCCTGTGACATACATGGAAATTCTCTGTACAAATTTTCTGTAAATCTAAAATTATTTTAGAATAGAAGGCTATTT
AAAAAAAAAAAGCTT

SEQ ID NO. 6: **NA10-1-A,B**

GACCGCTTGTGAATGCAAACAAAATTCAAATTTCCCTGAAAATTTATTCAACTTCTATATGCCAAGCACACCGCT
AAAGGCTTATCTTCTAAGTATATGCAGGCATACCCTACTCACACAAATAGCTTATTACCAGAGATAGGAAATTGC
AGGTAATTTGGGAGAAATTGTCATAGCCAAATTTATGGAAAAAATAAAATAAAAACTTCTCTATGGCCTCTTGAT
TTAAGAAAAAAACAGAACAATACTAAAAAAAAAAAGCTT

SEQ ID NO. 7: **NA11-4-A**

CAATCGCCGTTAGAATATACGTGACCACTGGTATTAGCTACTTCCTGCCAATAGGGGGCATTGTTTTGAGAAAA
ACAGCAGTCAGATTCGTCCCAGATGTCTACCTAAGGGTTCCTGGCAAAGGGGAGTCATTGTCCGAGACCTCAGTT
GCTTGCCTTTTTGGAATTTGATGGCCTCTAGGTGTGAGAAAGAAAAAAAACTTCCATAAGGTTAGATACGCAGGG
GTAAAACATGTATTATACTAGTAAAGAATTTAGTGCCAAAGATTTCAGAAATAAAAAGTGAAATATACTAATTAT
TCTAAAAAAAAAAAAGCTTAAGGGCGAATTC

SEQ ID NO. 8: **NA11-4-B,C**

GCCCTTAAGCTTTTTTTTTTTACATCACTTTAGAATATTTATTGTATTCCTTAATGCATTTCTTAACATGTATAG
CACTCTTTAATCAAGAATATAAAGTCATCTACTTAGAATCACATTATCTTAAAGATGCATACTGGAATGATAAGT
TTGAAGATGTAACTATCAACAATTCTTTTCAAAATCATATCAATATATTACTCTCATGGAACTTGCACATTCTAA

GAAGGGTCATTTTTTCCCCCCAGTACTGGGAAGGTATGCATTTAACCATGTGGTCAGCCAGAAAGGCTGTTTTAT
ATATGGTGTGTGTTACGGCGATTG

SEQ ID NO. 9: NA12-1-A,B,C
AAGCTTTTTTTTTTTACACTGGAAGGGTCCGATTGCTGGTAAATATGGCTCTATCTATCGCCGACTATCGCCCAC
TATCACCCACTATCGCCGAAGGGCGAATTC

SEQ ID NO. 10: NA15-1-A,L
TTCCGAACCCACTCCACCTTACTACCAGACAACCTTAACCAAACCATTTACCCAAATAAAGTATAGGCGATAGAA
ATTGAAACCTGGCGCAATAGATATAGTACCGCAAGGGAAAGATGAAAAATTATAACCAAGCATAATATAGCAAGG
ACTAACCCCTATACCTTCTGCATAATGAATTAACTAGAAATAACTTTGCAAGGAGAGCCAAAGCTAAGACCCCCG
AAACCAGACGAGCTACCTAAGAACAGCTAAAAGAGCACACCCGTCTATGTAGCAAAATAGTGGGAAGATTTATAG
GTAGAGGCGACAAACCTACCGAGCCTGGTGATAGCTGGTTGTCCAAGATAGAATCTTAGTTCAACTTTAAATTTG
CCCACAGAACCCTCTAAATCCCCTTGTAAATTTAACTGTAAAAAAAAAAAAGCTT

SEQ ID NO. 11: NA15-1-H
TTCCGAACCCATACTTTAGAGTACATACCAGTAAAATTCCAAATAAATTAAAATTTTAAATATAAAACATAAACC
ATATATGTAATGTGAATTAATTATTTTATATGCTTGGGGTAGTAAAGGGCTTTCATAATATGGTTTGAAATCCAG
ATGCCATGAAAGAGAAAATTAATACATTTTCTACACAAGAGTAAAACATTTCTGCATGGCAAAACGTGAAAGTAA
AGTCAAAACATAAATAACAAAGAGGTAAAAAACTTTTGTGCTTCATATCCAGATAGTAAATAATTTTTCTAATGT
AAAAAGAACTCAAATTACTTCATAGAAGACTAAAATATCAACAAAAAATTAGAGTAAGAATATCAACAGAGGGTT
CAAGGAAAAGTAAGTACAAAAACTTATACAATATTTTGAAGTGCTCAAACTCATTTATTAAAAAAAAAAAGCTT

SEQ ID NO. 12: NA16-3-Q
ANCCAGNGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCCTATTNTTCATANCCN
ANTACNCAAACATTATTANAATAAACACCCTCACCACTACAATCTTCCTANGAACAACATATGACGCACTCTCCC
CTNAACTCTACACAACATATTTNGTCACCAAGACCCTACTTCTAACCTCCCTGNTCTTAT

SEQ ID NO. 13: NA16-3-M
AAGCTTTTTTTTTTTAAGAGGAAAACCCGGTAATGATGTCGGGGTTGAGGGATAGGAGGAGAATGGGGGATAGGT
GTATGAACATGAGGGTGTTTTCTCGTGTGAATGAGGGTTTTATGTTGTTAATGTGGTGGGTGAGTGAGCCCCATT
GTGTTGTGGTAAATATGTAGAGGGAGTATAGGGCTGTGACTAGTATGTTGAGTCCTGTAAGTAGGAGAGTGATAT
TTGATCAGGAGAACGTGGTTACTAGCACAGAGAGTTCTCCCAGTAGGTTAATAGTGGGGGGTAAGGCGAGGTTAG
CGAGGCTTGCTAGAAGTCATCAAAAAGCTATTAGTGGGAGTAGAGTTTGAAGTCCTTGAGAGAGGATTATGATGC
GACTGTGAGTGCCGTTCGTAGTTTGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCTGATTCCGCCAAGCCCGTT
CCCTTGGCTGTGGTTTCGCTTGGCTAANGGCGAATTCCAGCACACTGGCGGCCCGTACTANTGGATCCCAAGCTC
GGTACCAAGCTTTGATGCATAGCTTGAGTATTCTATAGNGNCCCCTAATANCTTGGCCTAATCATGGCCATANCT
GGTTCCTGNGNGAAATTGGTATNCGNTCACAATTNCCCACAACNTCCGAA

SEQ ID NO. 14: NA16-3-P
GCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCCTGTTGAGCCTGAACTC
TACNCAACATATTTTGNCACCAAGACCCTACTTCTAACCTCCCTGTTCTTAT

SEQ ID NO. 15: NA16-3-L
GCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCCTGAACTCTACACAACA
TATTTTGTCACCAAGACCCTACTTCTAACCTCCCTGTTCTTATGAATTCGAACAGCATACCCCCGATTCCGCTAC
GACCAACTCATACACCTCCTATGAAAAAACTTCCTACCACTCACCCTAGCATTACTTATATGATATGTCTCCATA
CCCATTACAATCTCCAGCATTCCCCCTCAAACCTAAAAAAAAAAAAGCTT

SEQ ID NO. 16: NA16-3-O
AGCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCAACCGAACCCCCTTCG
ACCTTGCCGAAGGGGAGTCCGAACTAGTCTCAGGCTTCAACATCGAATACGCCGCAGGCCCCTTCGCCCTATTCT
TCATAGCCGAATACACAAACATTATTATAATAAACACCCTCACCACTACAATCTTCCTAGGAACAACATATGACG
CACTCTCCCCTGAACTCTACACAACATATTTTGTCACCAAGACCCTACTTCTAACCTCCCTGTTCTTATGAATTC
GAACAGCATACCCCCGATTCCGCTACGACCAACTCATACACCTCCTATGAAAAAACTTCCTACCACTCACCCTAG
CATTACTTATATGATATGTCTCCATACCCATTACAATCTCCAGCATTCCCCCTCAAACCTAAAAAAAAAAAAGCTT

SEQ ID NO. 17: NA16-3-K
AAGCTTTTTTTTTTTAATTAGAATTGTGAAGATGATAAGTGTAGAGGGAAGGTTAATGGTTGATATTGCTAGGGT
GGCGCTTCCAATTAGGTGCATGAGTAGGTGGCCTGCAGTAATGTTAGCGGTTAGGCGTACGGCCAGGGCTATTGG
TTGAATGAGTAGGCTGATGGTTTCGATAATAACTAGTATGGGGATAAGGGGTGTAGGTGTGCCTTGTGGTAAGAA

GTGGGCTGGGGCATTTTTAATCTTAGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCTGATTCCGCCAAGCCCGT
TCCCTTGGCTGTGGTTTCGCTGGCT

SEQ ID NO. 18: NA16-3-I

AAGCTTTTTTTTTTTAAGAGGAAAACCCGGTAATGATGTCGGGGTTGAGGGATAGGAGGAGAATGGGGGATAGGT
GTATGAACATGAGGGTGTTTTCTCGTGTGAATGAGGGTTTTATGTTGTTAATGTGGTGGGTGAGTCAAGCTCAAC
AGGGTCTTCTTTCCCCGCTGATTCCGCCAAGCCCGTTCCCTTGGCTGTGGTTTCGCTGGCT

SEQ ID NO. 19: NA16-3-N

AGCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAGAAGACCCTATTCTTCATAGCCG
AATACACAAACATTATTAGAATAAACACCCTCACCACTACAATCTTCCTAGGAACAACATATGACGCACTCTCCC
CTGAACTCTACACAACATATTTTGTCACCAAGACCCTACTTCTAACCTCCCTGTTCTTATGAATTCGAACAGCAT
ACCCCCGATTCCGCTACGACCAACTCATACACCTCCTATGAAAAAACTTCCTACCACTCACCCTAGCATTACTTA
TATGATATGTCTCCATACCCATTACAATCTCCAGCATTCCCCCTCAAACCTAAAAAAAAAAAAGCTT

SEQ ID NO. 20: NA16-3-R

AAGCTTTTTTTTTTTAAGAGGAAAACCCGGTAATGATGTCGGGGTTGAGGGATAGGAGGAGAATGGGGGATAGGT
GTATGAACATGAGGGTGTTTTCTCGTGTGAATGAGGGTTTTATGTTGTTAATGTGGTGGGTGAGTGAGCCCCATT
GTGTTGTGGTAAATATGTAGAGGGAGTATAGGGCTGTGACTAGTATGTTGAGTCCTGTAAGTAGGAGAGTGATAT
TTGATCAGGAGAACGTGGTTACTAGCACAGAGAGTTCTCCCAGTAGGTTAATAGTGGGGGGTAAGGCGAGGTTAG
CGAGGCTTGCTAGAAGTCATCAAAAAGCTATTAGTGGGAGTAGAGTTTGAAGTCCTTGAGAGAGGATTATGATGC
GACTGTGAGTGCGTTCGTAGTTTGAGTTTGCTAGACTAGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCTGATT
CCGCCAAGCCCGTTCCCTTGGCTGTGGTTTCCTGGCT

SEQ ID NO. 21: NA16-4-A

GACCGCTTGTACTGAAGGGAACAGAGACAGAATGAAATGAAAGAAGGCAGTTGAACTTCTAGGCTTCTACAGGCA
GAAAACAGGCTGATAGAACTGCTCAACTACAGACATGTTCTACCTTTCTAGAAAAAAAAAAAAGCTTAAGGGCGAA
TTC

SEQ ID NO. 22: NA16-4-Q

AANCTTTTGTTNTTTATGNGTTGGCNNGCAGGTNGAGGCTTACTAAAAGNGTGAAAACGTATGCTTGGATTAAGG
CTGACAGCGATTGCTAANGATAGTCAGTANAATTANAATTGTGAAGATGATAANTGTAGAGGGAAGGTTAATGGT
TGATATTGNTAGGGTGGCNCTNCNNNTTAGNTGCCNNACTANANTNAAGCTNAACAGGGTCTTCTTTCCCCNNTG
NTTCCGNCAAGCCCGTNCCCTTGGCTGNGGTTNCNCTGGCT

SEQ ID NO. 23: NA16-4-N

AAGCTTTTTTTTTTTTATAAGATTATTAGTATAAAAGGGGAGATAGGTAGGAGTAGCGTGGTAAGGGCGATGAGTG
TGGGGAGGAATGGGGTGGGTTTTGTATGTTCAAACTGTCATTTTATTTTTACGTTGTTAGATATGGGGAGTAGTG
TGATTGAGGTGGAGTAGATTAGGCGTAGGTAGAAGTAGAGGTTAAGCTCAACAGGGTCTTCTTTCCCCGCTGATT
CCGCCAAGCCCGTTCCCTTGGCTGTGGTTTCGCTGGCT

SEQ ID NO. 24: NA16-4-K

AAGCTTTTTTTTTTTAAGAGGAAAACCCGGTAATGATGTCGGGGTTGAGGGATAGGAGGAGAATGGGGGATAGGT
GTATGAACATGAGGGTGTTTTCTCGTGTGAATGAGGGTTTTATGTTGTTAATGTGGTGGGTGAGTGAGCCCCATT
GTGTTGTGGTAAATATGTANAGGGAGTATAGGGCTGTGACTAGTATGTTGAGTCCTGTAAGTAGGAGAGTGATAT
TTGATCAGGAGAACGTGGTTACTAGCACAGAGAGTTCTCCCAGTAGGTTAATAGTGGGGGGTAAGGCGAGGTTAG
CGAGGCTTGCTANAAGTCATCAAAAAGCTATTAGTGGGAGTAGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCT
GATTCCGNCAAGCCCGTTCCCTTGGCTGTGGTTTCNCTGGCT

SEQ ID NO. 25: NA16-4-1

ANNCNTNGNNNNNNNANCAANGGGAACGGGCTTGGNGGAATCAGCNGNGAAANAAGACCCTNANNTCTNAACANC
ATATTAANACACCAGAGACCCTACTTCTNACCTNCCTGGNCTTATGAATNNAANCAGCATACCCANNANTCCNCN
NCNACCAACTCATACNCCTCCTATGAAAAAACTNNCTACCACTCANCCTAGCATTACTTATATGATANGTCTCCA
TACCCNNTNNAATCTTCATNATTCCCNCTCTAACCTNANAAANNAAAGCTTAANGGCNAATNGNAACACACTGGC
GNCCNTTNCTANCGGANCCGAGCNNNNTACCNAGCTTGATGCATAGATTNNGTATTCTNTAGGGGTCACCTATAT
AGCTTGGNGTAATNTGGTCATAGCTGNNTCTGTGTNAAATGGCTANACGCTCACAATNNCACACNNTATACNAGC
NCGNANNNNTTNCGNNCNNAAGCCTGGCGTGCCTAATGAGTGAGCTAACTCACATTAATTNCCTTTNCNCCTCAC
TGNCCGCTNTCCNC

SEQ ID NO. 26: NA16-4-F

AACCTTTNANNNNTNANNANNNGNANCGGGCTNGGNNGANTNAGCNGNGNAANAAGANNATAAN
GNNNGANCGGCATATGAANATNAATCGACCCTANTAGGGCCTTCTTGNCCNNATGANTNCGACNA

GCNNNNNCCCTAGGCCGCTNCTACCATNGCTTACGNCNANTATNANNACACTGGCTACCNCTANTN
CTNGNNNNANTNANATGANNNNTCTNNATACCNATTNCGANCNTNNTNNNTCNNCCTCTAACNNN
CTAGGCNTAANCTTAAGGGCNAATGCACCTGTGTGANAGCCGTNTCTAGCTGGAACCNAGCNANN
NNCNANGNTNGATGNATATATNGAGTATTCTATAGNGGNGCCTAAAGAGCTAGCGCGTATCTNCA
TGGNATNCGTGCGCCTNCTGTGANANTGTTNATANCGNNAANAANTGTACAGNCNANTNNATNAC
GNGAANNNTNACAANNNNNGCCGAGGAGGNCCTAANGNGATATGCNNNCTCTTATGNTTNTCGGC
NCACNTNACTGNCNGCTNTCCGNGCCNGGNA

SEQ ID NO. 27: **NA16-4-H**

AAGCTTTTTTTTTTTAAGAGGAAAACCCGGTAATGATGTCGGGGTTGAGGGATAGGAGGAGAATGGGGGATAGGT
GTATGAACATGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCTGATTCCGCCAAGCCCGTTCCCTTGGCTGTGGT
TTCGCTGGCTAAGGGCGAATTC

SEQ ID NO. 28: **NA16-4-L**

AAGCTTTTTTTTTTTAAGAGGAAAACCCGGTAATGATGTCGGGGTTGAGGGATAGGAGGAGAATGGGGGATAGGT
GTATGAACATGAGGGTGTTTTCTCGTGTGAATGAGGGTTTTATGTTGTTAATGTGGTGGGTGAGTGAGCCCCATT
GTGTTGTGGTAAATATGTAGAGGGAGTATAGGGCTGTGACTAGTATGTTGAGTCCTGTAAGTAGGAGAGTGATAT
TTGATCAGGAGAACGTGGTTACTAGCACAGAGAGTTCTCCCAGTAGGTTAATAGTGGGGGGTAAGGCGAGGTTAG
CGAGGCTTGCTAGAAGTCATCAAAAAGCTATTAGTGGGAGTAGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCT
GATTCCGCCAAGCCCGTTCCCTTGGCTGTGGTTTCGCTGGCT

SEQ ID NO. 29: **NA16-4-E**

AAGCTTTTTTTTTTTATAAGATTATTAGTATAAAAGGGGAGATAGGTAGGAGTAGCGTGGTAAGGGCGATGAGTG
TGGGGAGGAATGGGGTGGGTTTTGTATGTTCAAACTGTCATTTTATTTTTACGTTGTTAGATATGGGGAG
TAGTGTGATTGAGGTGGAGTAGATTAGGCGTAGGTAGACTAGAGTCAAGCTCAACAGGGTCTTCTTTCCC
CGCTGATTCCGCCAAGCCCGTTCCCTTGGCTGTGGTTTCGCTGGCT

SEQ ID NO. 30: **NA16-5-A**

AAGCTTTTTTTTTTTTATAAGGGTGGAGAGGTTAAAGGAGCCACTTATTAGTAATGTTGATAGTAGAATGA
TGGCTAGGGTGACTTCATATGAGATTGTTTGGGCTACTGCTCGCAGTGCGCCGATTAGGGCGTAGTTTGA
GTTTGATGCTCACCCTGATCAGAGGATTGAGTAAACGGCTAGGCTAGAGTCAAGCTCAACAGGGTCTTCT
TTCCCCGCTGATTCCGCCAAGCCCGTTCCCTTGGCTGTGGTTTCGCTGGCTAAGGGCGAATTC

SEQ ID NO. 31: **NA16-5-G**

G	AATTCGCCCT	TAGCCAGCGA	AACCACAGCC		
101	AAGGGAACGG	GCTTGGCGGA	ATCAGCGGGA	GTCCGAACTA	GTCTCAGGCT
151	TCAACATCGA	ATACGCCGCA	GGCCCCTTCG	CCCTATTCTT	CATAGCCGAA
201	TACACAAACA	TTATTATAAT	AAACACCCTN	ACCACTACAA	TCTTCCTAGG
251	AACAACATAT	GACGCACTCT	NCCCTGAACT	CTACACAACA	TATTTTGNCA
301	CCAAGACCCT	ACTTCTAACC	TCCCTGGTCT	TAT**GAATTC**	

SEQ ID NO. 32: **NA17-1-D,E,F**

GAATTCGCC	CTTGACCGCT				
101	TGTGAATGCA	AACAAAATTC	AAATTTCCCT	GAAAATTTAT	TCAACTTCTA
151	TATGCCAAGC	ACACTGCTAA	AGGCTTATCT	TCTAAGTATA	TGCAGGCATA
201	CCCTACTCAC	ACAAATAGCT	TATTACCAGA	GATAGGAAAT	TGCAGGTAAT
251	TTGGGAGAAA	TTGTCATAGC	CAAATTTATG	GAAAAAATAA	AATAAAAACT
301	TCTCTATGGC	CTCTTGATTT	AAGAAAAAAA	CAGAACAATA	CTAAAAAAAA
351	AAAGCTTAAG	GGCGAATTC			

SEQ ID NO. 33: **NA19-1-A,B,C**

GAATTCG	CCCTTAAGCT				
101	TTTTTTTTTT	AAGATTGTTC	TAATTCTGGT	TGTAAACTGC	TATTTTAAAA
151	AACAAAACAA	ACAGAAAACA	TCAAAAACAC	AAAAAGATAT	TAAAACAGCA
201	AGTCTTTTGT	ACATCACTGT	AGCATAAGCT	GCTTGAGGTT	GTCATGCAGA
251	ATAGTATCCT	TCACGTCACG	GAAAACAAGG	CGGATGTTCT	CCGTGTTGAT
301	AGCAGTGGTG	AAGTGGTGGT	ATAAGGGCTT	CTGTTGCTGG	TCCCGACGTT
351	TGAAGGGCGA	ATTC			

SEQ ID NO. 34: NA19-2T-C

GAATTCG CCCTTAAGCT TTTTTTTTTT AAGATTGTTC TAATTCTGGT
151 TGTAAACTGC TATTTTAAAA AACAAAACAA ACAGAAAACA TCAAAAACAC
201 AAAAAGATAT TAAAACAGCA AGTCTTTTGT ACATCACTGT AGCATAAGCT
251 GCTTGAGGTT GTCATGCAGA ATAGTATCCT TCACGTCACG GAAAACAAGG
301 CGGATGTTCT CCGTGTTGAT AGCAGTGGTG AAGTGGTGGT ATAAGGGCTT
351 CTGTTGCTGG TCCCGACGTT TGAAGGGCGA ATTC

SEQ ID NO. 35: NA19-2T-A,F

G AATTCGCCCT TCAAACGTCG GGGCATTCCG
101 GATAGGCCGA GAAAGTGTTG TGGGAAGAAA GTTAGATTTA CGCCGATGAA
151 TATGATAGTG AAATGGATTT TGGCGTAGGT TTGGTCTAGG GTGTAGCCTG
201 AGAATAGGGG AAATCAGTGA ATGAAGCCTC CTATGATGGC AAATACAGCT
251 CCTATTGATA GGACATAGTG GAAGTGAGCT ACAACGTAGT ACGTGTCGTG
301 TAGTACGATG TCTAGTGATG AGTTTGCTAA TACAATGCCA GTCAGGCCAC
351 CTACGGTGAA AAGAAAGATG AATCCTAGGG CTCAGAGCAC TGCAGCAGAT
401 CATTTCATAT TAAAAAAAAA GCTTAAGGGC **GAATTC**

SEQ ID NO. 36: NA19-2T-B

GAATTCGC CCTTAAGCTT TTTTTTTTTA TATGAAATGA TCTGCTGCAG
151 TGCTCTGAGC CCTAGGATTC ATCTTTCTTT TCACCGTAGG TGGCCTGACT
201 GGCATTGTAT TAGCAAACTC ATCACTAGAC ATCGTACTAC ACGACACGTA
251 CTACGTTGTA GCTCACTTCC ACTATGTCCT ATCAATAGGA GCTGTATTTG
301 CCATCATAGG AGGCTTCATT CACTGATTTC CCCTATTCTC AGGCTACACC
351 CTAGACCAAA CCTACGCCAA AATCCATTTC ACTATCATAT TCATCGGCGT
401 AAATCTAACT TTCTTCCCAC AACACTTTCT CGGCCTATCC GGAATGCCCC
451 GACGTTTGAA GGGC**GAATTC**

SEQ ID NO. 37: NA19-2b-A,B,C

GAATTCGCC CTTAAGCTTT TTTTTTTTAA GATTGTTCTA
151 ATTCTGGTTG TAAACTGCTA TTTTAAAAAA CAAAACAAAC AGAAAACATC
201 AAAAACACAA AAAGATATTA AAACAGCAAG TCTTTTGTAC ATCACTGTAG
251 CATAAGCTGC TTGAGGTTGT CATGCAGAAT AGTATCCTTC ACGTCACGGA
301 AAACAAGGCG GATGTTCTCC GTGTTGATAG CAGTGGTGAA GTGGTGGTAT
351 AAGGGCTTCT GTTGCTGGTC CCGACGTTTG AAGGGCGAAT TC

SEQ ID NO. 38: NA19-3-A,C

GAATTC
101 GCCCTTCAAA CGTCGGGGCA TTCCGGATAG GCCGAGAAAG TGTTGTGGGA
151 AGAAAGTTAG ATTTACGCCG ATGAATATGA TAGTGAAATG GATTTTGGCG
201 TAGGTTTGGT ATAGGGTGTA GCCTGAGAAT AGGGGAAATC AGTGAATGAA
251 GCCTCCTATG ATGGCAAATA CAGCTCCTAT TGATAGGACA TAGTGGAAGT
301 GAGCTACAAC GTAGTACGTG TCGTGTAGTA CGATGTCTAG TGATGAGTTT
351 GCTAAAAAAA AAAAGCTTAA GGGCGAATTC

SEQ ID NO. 39: NA19-3-B

GA ATTCGCCCTT AAGCTTTTTT
101 TTTTTAGAAT TAAGATTGTT CTAATTCTGG TTGTAAACTG CTATTTTAAA
151 AAACAAAACA AACAGAAAAC ATCAAAAACA CAAAAAGATA TTAAAACAGC
201 AAGTCTTTTG TACATCACTG TAGCATAAGC TGCTTGAGGT TGTCATGCAG
251 AATAGTATCC TTCACGTCAC GGAAAACAAG GCGGATGTTC TCCGTGTTGA
301 TAGCAGTGGT GAAGTGGTGG TATAAGGGCT TCTGTTGCTG GTCCCGACGT
351 TTGAAGGGCG AATTC

SEQ ID NO. 40: NA22-3-B

GA ATTCGCCCTT AAGCTTTTTT
101 TTTTTACTCT CAGGTTCAGG GTACTAAGTT GAAGTTCTTA CTAGGAAAGA
151 TGCATATTAA TAATGTATTT GTGGCTTCTT GAGTGCACAG AAGTGATTCT
201 GACATATGGG CAGGAAAAGT GACATTCAGG TGAAAACACT ATGGCCAGGG
251 ATCAAAGGGC **GAATTC**

SEQ ID NO. 41: NA22-3-D

GAATTCG CCCTTAAGCT
101 TTTTTTTTTT AAGATAAATG TTGAATTGCA GGAAGAATAA CATTTTGGAA
151 CAGTAATGTG GGATATAAGA GAAAGTCACA TAGCTCCAAA TTTAGGGTGA
201 GACTTTACAT GTCTTANAAG ACCATTAAGA GGACTTCCAA CAAGTAGGGG
251 AGACCAAGTT TCAATTAGGG CAGAAGATAG GGAAGGAACT CTATAAAGAG
301 ACTAAAACTG TGAGGGTTCG CTGGCTAAGG GC**GAATTC**

SEQ ID NO. 42: NA22-3-F

GAATTCGC CCTTTGATCC
101 CTGGCACTTG AACACTAATG AATATTATGA CTGCCACTTT AAAGGAGGCA
151 GAAGAAGTTT AAAAAGTAAA ACAAAAAGTT TGTTTCAGAA AACAAGCATT
201 TTACCTCTGT TTCAAATAGT CTAATTTTTT TAGTGATGAA AACTTCTGAG
251 ACCAGTAGAT GTTTGTAAAT AAAAAACATT TATGGCAGTC TTTGTAACTG
301 TAATGAAACT GGTAGAGAGT AATAATAGCC TGTTTTTTGT TTGTTTGTTT
351 GCTTTGGGGG ATATTTGCAA TACAGTTTAT TGATATGTCA CATACATGTA
401 AAGTTTATAA TTCACCCCAG AATTTATATT ACTAAGTTTG TGCTAGTATT
451 AAAGAGCTTT TCAAATTCAG TGCCTGTTTA AAAAAAAAAA GCTTAAGGGC
501 GAATTC

SEQ ID NO. 43: NA22-3-G

GAATTCG CCCTTAAGCT
101 TTTTTTTTTT ATGTGTTGTC GTGCAGGTAG AGGCTTACTA GAAGTGTGAA
151 AACGTAGGCT TGGATTAAGG CGACAGCGAT TTCTAGGATA GTCAGTAGAA
201 TTAGAATTGT GAAGATGATA AGTGTAGAGG GAAGGTTAAT GGTTGCCAGG
251 GATCAAAGGG C**GAATTC**

SEQ ID NO. 44: NA22-3-C,E

GAATTCGCCC TTTGATCCCT GGATAGAAAG CCTGAGCCCA TTGGATCTGT
151 GAAAGCCTCT AGCTTCACTG GTGCAGAAAA TTTTCCTCTA GATCAGAATC
201 TTCAAGAATC AGTTAGGTTC CTCACTGCAA GAAATAAAAT GTCAGGCAGT
251 GAATGAATTA TATTTTCAGA AGTAAAGCAA AGAAGCTATA ACATGTCGTG
301 TACAGTACAC TCTGAAAAGA AATCTGAAAC AAGTTATTGT AATGATAAAA
351 ATAATGCACA GGCATGGTTA CTTAATATTT TCTAACAGGA AAAGTCATCC
401 CTATTTCCTT GTTTTACTGC ACTTAATATT ATTTGGTTGA ATTTGTTCAG
451 TATAAGCTCG TTCTTGTGCA AAATTAAATA AATATTTCTC TTACCTTATA
501 AAAAAAAAAA GCTTAAGGGC **GAATTC**

SEQ ID NO. 45: NA26-3-A

GAATTCG
101 CCCTTAAGCT TTTTTTTTTT ACAGATGTGC AGGAATGCTA GGTGTGGTTG
151 GTTGATGCCG ATTGTAACTA TTATGAGTCC TAGTTGACTT GAAGTGGAGA
201 AGGCTACGAT TTTTTTGATG TCATTTTGTG TAAGGGCGCA GACTGCTGCG
251 AACAGAGTGG TGATAGCGCC TAAGCATAGT GTTAGAGTTT GGATTAGTGG
301 GCTATTTTCT GCTAGGGGGT GGAAGCGGAT GAGTAAGAAG ATTCCTGCTA
351 CAACTATAGT GCTTGAGTGG AGTAGGGCAG AGCAAGGGCG AATTC

SEQ ID NO. 46: NA26-3-B,C

GAA TTCGCCCTTT GCTCTGCCCT
101 AATTGAATTT GCAACTGTTA ATTACTTCAC CAAAAGAGGA TGGGCTTGGG
151 ATGGGAAGAG TGTAGTAAAT GACAAGTCCC CTTCAATAAA AGCTGAAGGC
201 ATTACATTAA CATACAACTC AGTGAAAGCA ATTCTTCAAG GAGCTAAGCT
251 AATATGGTCC AAGTATATAG CTTTCTCATG GCCCAGTTTA TTCCAAGAAA
301 AGACTTTAGA GTACTTAGAG AAGTGGATGG ACTGTTTAAC CTTCAAACAA
351 TCTTCTAAAA AAAAAAAGCT TAAGGGCGAA TTC

SEQ ID NO. 47: NC4-1-F

GAATTCGC CCTTAATCGG GCTGGAGCTA TTGATTAGCA AGTAAGTAGG
151 CGTTTGCTAA AACTAGAGAG AGAATTTATG AGGTTATTCA GGGAGAGGAT
201 ATAGGGTGAT AATTACAATG GACAAAGAAT AGATCTTGAG CTGCACAAAC
251 ATTTAAGGCA CAGGTAGAAG AAAAGGAGTC TATGTTAAGA GAAGGAATGG

301 TCAGAGAAAC AAGAGGGGGA ACTAGGAGAA AATGGTATTA TGAAAAACAA
351 AGGAGTAGAA ATTTGAAAAA AAAAAAGCTT AAGGGCGAAT TC

SEQ ID NO. 48: NC4-1-G,H

GAATTC GCCCTTAAGC TTTTTTTTTT TCAACAGCAA CACAGGTTTA
151 TTACNAGCAA AACCCTGCGG AGGGGGAAAC CAGCTTAGTG TCAGTGCCCA
201 CTGCCGCTCA CAGGCTGGGG TAATCATAGC GCTGGGAGGG AGGGCTCTGG
251 ACAGTATAGC TTGCTGCTCA GTAGAAGATG ATAAGGATGT TCCTGAAGTC
301 AGGCTGTTGG GCCTTTGCCC AGCAGGATGT GATAAGGATG TTTCTGCAGT
351 CAGGTGGTTA GGACATTTCT CACAGCCCGA TTAAGGGCGA ATTC

SEQ ID NO. 49: NC16-4-G,C,O

GAATTCG CCCTTAAGCT
101 TTTTTTTTTT CATCCAGTTT TGAAGTAACA TCTTCCTTCC GAACAATCAC
151 CTGCTTTATT GATGGACGTT CTGTTTCTTT GAATCTTTGA GATCTATATG
201 CATCAATGCT GTAAAGAAGA TCACGATCTT CAGAACCAAG GCTATCACCA
251 GATTCAGCTC GAGGGACACG AGTTCTTTGG AATTTTCCTG GTTTTGGACT
301 TTCATCACTT CTGCTGGTGT CTTTCAATTC CAGTCTAGGT GTGGACACTA
351 AACTCTCTGG ACTTACCACA CCAACTGTTT GTGCCAATGG TGCAAGTAAA
401 GACATTGAGG AGATATTCAG TGCATCTTCC TGTTCTTCGC TGGCTAAGGG
451 CGAATTC

SEQ ID NO. 50: NC17-1-A,H

GAATTCGCC CTTGACCGCT TGTGAATAAT ATTGTCTCTA
151 TAGGTGTGCA AGCATTTCCT GGAAGCTATT GAAAACAACA AGTATGGCTG
201 GTTTTGGGTA TGCCCTGGAG GGGGTGATAT TTGCATGTAT CGTCATGCAC
251 TTCCTCCTGG ATTTGTGTTG AAAAAAAAAA AGCTTAAGGG C**GAATTC**

SEQ ID NO. 51: NC17-1-D,G

GA ATTCGCCCTT GACCGCTTGT
101 ACTGAAGGGA ACAGAGACAG AATGAAATGA AAGAAGGCAG TTGAACTTCT
151 AGGCTTCTAC AGGCAGAAAA CAGGCTGATA GAACTGCTCA ACTACAGACA
201 TGTTCTACCT TTCTAGAAAA AAAAAAAGCT TAAGGGCGAA TTC

SEQ ID NO. 52: NC17-2-A

GAAT TCGCCCTTGA CCGCTTGTTA
101 AAAGGAAAAA AGTTGAGAAG TCAGGCCTTG AAAAGAGGAT AGACCAGGCT
151 GTGGAGGAGT GGAATATTGA GAAGGCTGAG GAACTCAGCA ACCAGCTAGC
201 TACTCGAGAG GAAAATCATG GGAGTAATGT TGTGTGTTTC TCAGTGTCAG
251 CGATCCAGAG ACTCGTGCTG TCTCTTTGTC CCTTTCCTGA TGGTGTTACT
301 TCGATTAACT ACTTCAGCTT GGTGTAAAAA TTGCCAAAGC AGTTGCCTGC
351 CACAACTTTG TAAAAGCCAA AAAGGAGGTT GAAAATTCAC AGGCTGCCCG
401 AAAAAAAAAA AGCTTAAGGG C**GAATTC**

SEQ ID NO. 53: NC17-2-C,F,H

GA ATTCGCCCTT GACCGNTTGT
101 ACTGAAGGGA ACAGAGACAG AATGAAATGA AAGAAGGCAG TTGAACTTCT
151 AGGCTTCTAC AGGCAGAAAA CAGGCTGATA GAACTGCTCA ACTACAGACA
201 TGTTCTACCT TTCTAGAAAA AAAAAAAGCT TAAGGGCNAA TTC

SEQ ID NO. 54: NC17-4-A

GAATTCG
101 CCCTTGACCG CTTGTTAAGA GGAACTGATC TCATATATTT GTATCAGAAC
151 TGTATTTTTA TGTTATATTG TATAGTTTGC TCTCCTGCCC CTCTCCTTAA
201 AACTGAATGG TGCCAATAAT TTGATACTAA TGACTACAAA AAAAGGTAAT
251 GCCTCATTTA CTAGTATTGT TGTAAAATGA GGAATGTATG TGAATATTCA
301 GATAACCGAG GATTAACCCT TTAAGTGCTG AATCTTTAAA ATTTTAATAT
351 ATTTTTTTTG AGGGAAATCT TTCTAAAATG TATTACGCAC TTCCCTGCCT
401 TAGTAAACAG AGTATACTGG AAAAAAAAAA AGCTTAAGGG C**GAATTC**

SEQ ID NO. 55: NC17-4-H

GA ATTCGCCCTT GACCGCTTGT

```
101  GGATGGAGAA  GGGGAGAGCA  TCTAGGCAGG  CAAACAGAAG  GGAAGTGGAG
151  TTAAACCTCT  GGCATGAAGT  CTGGGAGTAG  GGTAGGCTAG  GGGGTTTCTT
201  CTATGACACT  TGACCCTTCC  ATGCTGGTTC  CCAAGCCTAT  TGGAGGAATG
251  TGGGTGTGGC  CGAGGTGATG  GCAAGAAAGG  TGCAAGAAAG  TGAGCAGTCT
301  GCCTGTGAGT  GAGCACAGAT  GCCGGGGTGT  GTGTGTGTGT  GTGTGATTTT
351  CACTGTGGGG  TGTGTCTGTG  AGAGCTAGCT  GCCTTACCCC  TCCTTGGCAC
401  ATAGTAGGCC  TTCCATAAAT  GTTGGATGGA  TGGATAAATA  GATTGGGACC
451  ATCAGACCAT  GAAAAAAAAA  AAGCTTAAGG  GCGAATTC
```

SEQ ID NO. 56: NC17-4-E,F

```
GAATTC
101  GCCCTTGACC  GCTTGTGATG  AAACTGTAAC  TTACAAGAAA  AGGGCTGGGT
151  TTTGAAAATA  ACACAGGCTC  TAAAAACCCT  AAGAAGCGGT  GCAACTTTTG
201  GCAGGAATCG  GGGTTAGCGG  GACCTCAAGG  GCTCACTGCG  GCTAAGTGAA
251  CGCTGACTGG  TCCTCCAGCG  TGAGCTAGAA  CAGACGTCTC  TATGGTCAAG
301  TAAACAGAGC  GTGTGCTGTC  TTCCCCATGT  GGTGGGGTTG  CGCATGATCA
351  GTAGCTGCAC  CACTAGAAAG  ATGGCGGAGC  AAGAGCAAAG  AAGAATCCCT
401  TTGGTTCCAG  AAAATCTCCT  GAAAAAAAAA  AAGCTTAAGG  GCGAATTC
```

SEQ ID NO. 57: NC17-5-A,C

```
GAATTC
101  GCCCTTGACC  GCTTGTACTG  AAGGGAACAG  AGACAGAATG  AAATGAAAGA
151  AGGCAGTTGA  ACTTCTAGGC  TTCTACAGGC  AGAAAACAGG  CTGATAGAAC
201  TGCTCAACTA  CAGACATGTT  CTACCTTTCT  AGAAAAAAAA  AAAGCTTAAG
251  GGCGAATTC
```

SEQ ID NO. 58: NC17-5-D

```
GAATTCGCCCTTGACCGCTTGTGAGGAGGAAAGTAATGCTGGGAAACTTGATATGTGTAAATAGAAAATATATAA
GCAAAGTTATCAGCCAGTCTTGATGTTGCAGCGGAAGTTGAGAGTGCCGTGGTATATCCTGTTTTGTGCATTAGC
TTTTTCTGGGGCATGAGCATTCAGGCATTTTATGAAGAACTTAGAAAAAGTGAAAAATATTTTGAAGTTTTATAT
TTTTGATCATTAGCTGGAAGGTTTGTCCAGTAGTAAGTTACTTGTGAGGTTTATAAAATATTAGGAACATTTGGC
AAGAAGAGACAGGTTTTGTGGGAATAATTTGTTACCTGTTGACCCTCACTGTGGACATATTTGTGTGTGTGTACC
TGTGTGTGTGTGTGTGTGTGTGTGTAAAAAGGAGGGTTTATAAAAAAAAAAAGCTTAAGGGCGAATTC
```

SEQ ID NO. 59: NC17-6-C

```
GAATTCG  CCCTTGACCG
101  CTTGTATTAT  CAGTGAATAT  AAATGTACTA  CATTTGCATG  CCTTTTGGGT
151  TTGCCTTAAT  TCTTACCTCA  TTTGCATCCT  ATCGATCTGG  AAAGAGCTGT
201  TTTGGATGAA  TGCAGTATAA  AATGTAAAAA  CCCTGCTAAA  TGACTTATTG
251  ATTAAGTATA  TCTATCTATA  TATACATATA  CACAAAGATA  TTATTTATCG
301  AAAGTAAAAA  AGATGGAAGT  GTATTGGTTT  CTGTTTGAAT  TTTCAAAGGC
351  TTCCAATGTG  GTGGCAATAA  ATGTCCCAAA  TAAATTTATA  ACAATTGAAA
401  AAAAAAAAGC  TTAAGGGCGA  ATTC
```

SEQ ID NO. 60: NC17-6-F

```
GAATTCG  CCCTTGACCG
101  CTTGTCAGAA  GATGAACATG  TATAGTGGCT  AACTTAAGGG  GAGTGGGTGA
151  CCCTGACACT  TCCAGGCACT  GTGCCCAGGG  TTTGGGTTTT  AAATTATTGA
201  CTTTGTACAG  TCTGCTTGTG  GGCTCTGAAA  GCTGGGGTGG  GGCCAGAGCC
251  TGAGCGTTTA  ATTTATTCAG  TACCTGTGTT  TGTGTGAATG  CGGTGTGTGC
301  AGGCATCGCA  GATGGGGGTT  CTTTCAGTTC  AAAAGTGAGA  TGTCTGGAGA
351  TCATATTTTT  TTATACAGGT  ATTTCAATTA  AAATGTTTTT  GTACATAGTG
401  AAAAAAAAAA  AGCTTAAGGG  CGAATTC
```

SEQ ID NO. 61: NC17-6-D,E,G

```
GA  ATTCGCCCTT  GACCGCTTGT
101  ACTGAAGGGA  ACAGAGACAG  AATGAAATGA  AAGAAGGCAG  TTGAACTTCT
151  AGGCTTCTAC  AGGCAGAAAA  CAGGCTGATA  GAACTGCTCA  ACTACAGACA
201  TGTTCTACCT  TTCTAGAAAA  AAAAAAAGCT  TAAGGGCGAA  TTC
```

SEQ ID NO. 62: NC17-7-A,B,C

```
GAATTCG  CCCTTGACCG
```

```
101  CTTGTACTTC  GAATCTATTT  TTGAAGTCGT  ATTCTCACAG  CATTCATGCT
151  TCACAGATGG  ACAGATGGAT  CCACTTGAGC  ACTTTTCTTT  GATAAATTGG
201  ACTAATTTAT  CTTAATAATA  TGAGGACACC  ATCTAAAGGA  ACTTTATAAT
251  TTATCATAAT  AAGGAGGTAA  CCATACAATA  TTTAAAAGAA  AATGAATCCT
301  TTTTTTATTT  TAAAGCTCAT  TGTTCTGAAT  GAAATACTAC  AGACCTGTAT
351  TGTAAACAAA  AAGAAAATGG  GGAAAAAAAA  AAAGCTTAAG  GGCGAATTC
```

SEQ ID NO. 63: **NC17-8-A,C,D**

```
GAATT
101  CGCCCTTGAC  CGCTTGTACT  GAAGGGAACA  GAGACAGAAT  GAAATGAAAG
151  AAGGCAGTTG  AACTTCTAGG  CTTCTACAGG  CAGAAAACAG  GCTGATAGAA
201  CTGCTCAACT  ACAGACATGT  TCTACCTTTC  TAGAAAAAAA  AAAAGCTTAA
251  GGGC
```
GAATTC

SEQ ID NO. 64: **NC17-8-E**

```
G  AATTCGCCCT  TGACCGCTTG  TATAATATAT
101  GTTCCCAGGC  AAGAAAATTT  TCGTGGTATC  AAAGCAAAGT  GGAAATCAGA
151  AAATGTGAAG  GTAGTCTAAA  TGTCTTGCAA  GCAGAAGTTT  GGTAGGACCA
201  GACATACGAT  TTAGTTAATG  GTCTATTACT  TTCCACTGAA  AAGCTTGTTT
251  TATATTAAAA  ATGGATCATT  TCATTTGAAG  TACAGTTGGT  CCTCTGTATT
301  CATGGGTTCT  GCAGCCAACG  ATTCAACCAA  CATGGATGGA  AAATATTTGA
351  AAAAAAAAAA  GCTTAAGGGC  GAATTC
```

SEQ ID NO. 65: **NC17-8-H,I**

```
GA  ATTCGCCCTT  AAGCTTTTTT
101  TTTTTCTGAT  TAAGTTACAA  ACATTCTCCC  TATAGCTAAA  CTCCGTGACT
151  AGGCTCCCAG  CCTCATGGCC  AAGAACAATA  AGTTCACCCA  CTTATCTGGA
201  GTAACCATAC  TAGATTAAAG  AAATACAATT  CTTTCTTCTA  AAGACAATTT
251  CCAGAAAGAC  CTGCCTTTCC  CTATGGGTAC  TTGACACTAG  GTCCCAGCAC
301  AGGCTAATCG  CTGTATGGTT  TCTTCGAAGA  TTGGCTTTTC  TCAGTTTCTT
351  TCTCTTTGAT  ACTGTACAAG  CGGTCAAGGG  CGAATTC
```

SEQ ID NO. 66: **NC17-9-B,F,G**

```
GAATTC
101  GCCCTTGACC  GCTTGTTAAA  ATATTTAAGT  ACCAGTTAAC  TAGCCAGCCA
151  ACATGGAACG  GGTATAAAGA  CCCAGTCTCT  GCCTTGAAGA  CCTACCATCT
201  AGCAGATGGA  GAGGGACATG  CTAACAAATA  GGGGCGCTAA  GTTTTTAGAC
251  TGCTATGACA  GAAGATTTAA  CAAAGGACAG  TGGGAGAACA  AAAAGAAGGG
301  GTTAAATCTA  CCTGGTGGTG  GAGTATGTCA  GGAAAGACTT  CTTCAGATTG
351  GCAATTTGGC  CTGAATCTAG  AAAAAAAAAA  AGCTTAAGGG  CGAATTC
```

SEQ ID NO. 67: **NC17-9-C,E**

```
GAATTCGCC  CTTGACCGCT  TGTCCAGGAA  GGGTTCCATC  AATGGTGAGC
151  ACCAGCCTGA  ATGCAGAAGC  GCTCCAGTAT  CTCCAAGGGT  ACCTTCAGGC
201  AGCCAGTGTG  ACACTGCTTT  AAACTGCATT  TTTCTAATGG  GCTAAACCCA
251  GATGGTTTCC  TAGGAAATCA  CAGGCTTCTG  AGCACAGCTG  CATTAAAACA
301  AAGGAAGTTC  TCCTTTTGAA  CTTGTCACGA  ATTCCATCTT  GTAAAGGATA
351  TTAAATGTTG  CTTTAACCTG  AACCTTGAAA  AAAAAAAAGC  TTAAGGGCGA
401  ATTC
```

SEQ ID NO. 68: **NC17-10-A,H**

```
GAATTC
101  GCCCTTGACC  GCTTGTACTG  AAGGGAACAG  AGACAGAATG  AAATGAAAGA
151  AGGCAGTTGA  ACTTCTAGGC  TTCTACAGGC  AGAAAACAGG  CTGATAGAAC
201  TGCTCAACTA  CAGACATGTT  CTACCTTTCT  AGAAAAAAAA  AAAGCTTAAG
251  GGC
```
GAATTC

SEQ ID NO. 69: **NC17-10-B,C,D**

```
GAATTCGCC  CTTGACCGCT
101  TGTTGACAGG  ATATGGGAGA  TGGAAAAGGA  AAGGATCTGC  ATCTAGTGAT
151  TGGAAATATA  GGAGTGGTGG  GGGTTAGTTT  CAGATGCCTG  TGGGATATTT
201  AATGTCCTGT  GTTGAGTTGG  AACTATGAGT  TCTACAGAGG  GCAAGATTTA
```

251 GGAGTTGGCA CTCCTAAGTG TCAATACATG TGAATAGGAT CGCTTTGGAG
301 GGTGAGAAGA GGTCTGAGAA CACTACTAGG GAACAGTGAA GGAAAAAAAA
351 AAAGCTTAAG GGC**GAATTC**

SEQ ID NO. 70: NC17-11-A,D,F
GACCGCTTGTACTGAAGGGAACAGAGACAGAATGAAATGAAAGAAGGCAGTTGAACTTCTAGGCTTCTACAGG
CAGAAAACAGGCTGATAGAACTGCTCAACTACAGACATGTTCTACCTTTCTAGAAAAAAAAAAAGCTT

SEQ ID NO. 71: NC17-11-E,H
GAATTCGCCCTTAAGCTTTTTTTTTTTCTATCTGAGGGGGGTCATCCGTAGGGACGAGAAGGGATTT
GACTGTAATGTGCTATGTACGGTGAATGGCTTTATGTACTATGTACTGTTAAGGGTGGGTAGGTTTGTTG
GTATCCTAGTGGGTGAGGGGTGGCTTTGGAGTTGCAGTTGATGTGTGATAGTTGAGGGTTGATTGCTGTA
CTTGCTTGTAAGCATGGGGAGGGGGTTTTGATGTGGATTGGGTTTTTATGTACTACAAGCGGTC

SEQ ID NO. 72: NC19-1-D
GAATTCGCCCTTCAAACGTCGGAGCATGGGCATGGTGAATGGCTTCTAGCTGTTGAAGAATGAAGTC
AAAAGAATGTATTTGGGGATGGAATAGCTGCAATTTGAGTTCATAACTTTTCTTTAGTTTCATTTTTGCG
GTCATGTCCCTGTATCCCTGAGGATGAAAACTGGAGATAACTCTTTACAAGCTCAAATGCTTAGATAAGG
GTGAGTTATAAAAAAGATATTTCTGCTACAGGAGAAGTAGTATTCATGTTTAATCTGGTCGGACATCACC
TGTTTTTCCCTTGGGTGACTTTGCTTGAAAAAAAAAAAGCTT

SEQ ID NO. 73: NC19-2-A,B
GAATTCGCCCTTCAAACGTCGGGGGAACATCAGGGGAACAAAACTGGAGAAAGATGCAGGGGGAAGGA
GAGTAGGAGAAAAGGGAGGAAGAAGAGAGAGAGATAATATGATTTGCTTTAAAAACAATTGCCTTTGT
TTAATACTCAGTAAAAGTTCAGAGTTCTTATTCTAAGTTGA**GAATTC**

SEQ ID NO. 74: NC19-2-C,E,F,G
GAATTCGCCCTTCAAACGTCGGGGCATTCCGGATAGGCCGAGAAAGTGTTGTGGGAAGAAAGTTAGA
TTTACGCCGATGAATATGATAGTGAAATGGATTTTGGCGTAGGTTTGGTCTAGGGTGTAGCCTGAGAATA
GGGGAAATCAGTGAATGAAGCCTCCTATGATGGCAAATAGAGCTCCTATTGATAGGACATAGTGGAAGTG
GGCTACAACGTAGTACGTGTCGTGTAGTACGATGTCTAGTGATGAGTTTGCNAAAAAAAAAAAAGCTT

SEQ ID NO. 75: NC19-2-D,H
GAATTCGCCCTTAAGCTTTTTTTTTTCAGATTGTTCTAATTCTGGTTGTAAACTGCTATTTTAAAAAACAAAACA
AACAGAAAACATCAAAAACACAAAAAGATATTAAAACAGCAAGTCTTTTGTACATCACTGTAGCATAAGCTGCTT
GAGGTTGTCATGCAGAATAGTATCCTTCACGTCACGGAAAACAAGGCGGATGTTCTCCGTGTTGATAGCAGTGGT
GAAGTGGTGGTATAAGGGCTTCTGTTGCTGGTCCCGACGTTTG

SEQ ID NO. 76: NC19-3-A,B,C
GAATTCGCCCTTCAAACGTCGGGGCATTCCGGATAGGCCGAGAAAGTGTTGTGGGAAGAAAGTTAGATTTACGCC
GATGAATATGATAGTGAAATGGATTTTGGCGTAGGTTTGGTCTAGGGTGTAGCCTGAGAATAGGGGAAATCAGTG
AATGAAGCCTCCTATGATGGCAAATACAGCTCCTATTGATAGGACATAGTGGAAGTGAGCTACAACGTAGTACGT
GTCGTGTAGTACGATGTCTAGTGATGAGTTTGAAAAAAAAAAAGCTTAAGGGCGAATTC

SEQ ID NO. 77: **NC19-4-A,B,C**

GAATTCGCCCTTCAAACGTCGGGGCATTCCGGATAGGCCGAGAAAGTGTTGTGGGAAGAAAGTTAGATTTACGCC
GATGAATATGATAGTGAAATGGATTTTGGCGTAGGTTTGGTCTAGGGTGTAGCCTGAGAATAGGGGAAAT
CAGTGAATGAAGCCTCCTATGATGGCAAATACAGCTCCTATTGATAGGACATAGTGGAAGTGAGCTACAA
CGTAGTACGTGTCGAAAAAAAAAAAGCTT

SEQ ID NO. 78: **NC19-5-B**

GAATTCGCCCTTCAAACGTCGGGGCATTCCGGATAGGCTGAGAAAGTGTTGTGGGAAGAAAGTTAGATTT
ACGCCGATGAATATGATAGTGAAATGGATTTTGGCGTAGGTTTGGTCTAGGGTGTAGCCTGAGAATAGGG
GAAATCAGTGAATGAAGCCTCCTATGATGGCAAATACAGCTCCTATTGATAGGACATAGTGGAAGTGAGC
TACAACGTAGTACGTGAAAAAAAAAAAGCTT

SEQ ID NO. 79: **NC19-5-D**

GAATTCGCCCTTCAAACGTCGGCAGGAACTTGCTCGACTGAGAGACTCAGCCTCCAGAGTAGTTGGG
ATTACAGACACGCACCACCGCGCCCGGCCATCATGACTTTTCTCTGCTTCTTGAGAGCACTTCCAGCATC
GCTAGTCGCACTTTGTGACTCTCACAGAAGGAGGAAGAGGAGGACACTTTTATTGAAGAACAACAACTAG
AAGAAGAGAAGCTATTGGAAAAAAAAAAAGCTT

SEQ ID NO. 80: **NC24-1-A,B,C**

GAATTCGCCCTTAANCTTTTNTTNTTTCAAGANGAGCTGTNTNGNTANNATGCTNAGCTGTNTGATAGGNCTNAC
CANGTCATANNTTNAGGTTNGCCATGGNCNNACTACTNGGACCCAACATGAAATATGACNNNCNNTTNGG
CATAAAAGAGGCACACGGGAACATCTGATGGANTAAAAAATAACTATTATTAATGCNACTACTAATATGA
ATATCTTATTACACAAACAGGAAGAATTACGTATTTTACAGGGTATTGGTGAGCAGTCAAAAAGCGTGGC
AAATTACCTAAAANGTTTANAAGGTTTAAGTGATCAAATATTTGCATNANATATAATTNCCCCCNNTAAA
GAACTTTGTATTTAAATGTGTTTTACTATAAGCACAGAATTAACCTTTGCTCTCCTGNANGTACCCCANN
TTTGNNCATACAGAAGANGCATGGGCCTATCTCATACGTATGCNCATACNAACACACATTCACAAACANG
GAAAAAACGAATGCTAAAAGTCTAAAAGTACTCCANNCNNANGGCGAATTC

SEQ ID NO. 81: **NC26-1-A,B,C**

GAATTCGCCCTTTGCTCTGCCCTACTTAATCACTAACACATCTTATACTGTCTAACCTCCAGAATTT
TGTTGAGATTCTCTGCTGATGTGTTTCCTTTGTTTCCTGTTCTCTCTATCACTTAGAGTTTGTTGTATTT
AATACCTTTGCTATCATTTTATTGTGGTTTTGGTTGGGAGAGGAAATAAAATGGCCAATCCACTACCTCG
AAAAAAAAAAAGCTTAAGGGCGAATTC

SEQ ID NO. 82: **NC26-1*-A,B,C**

TGCTCTGCCCTACTTAATCACTAACACATCTTATACTGTCTAACCTCCAGAATTTTGTTGAGATTCTCTGCTGAT
G TGTTTCCTTTGTTTCCTGTTCTCTCTATCACTTAGAGTTTGTTGTATTTAATACCTTTGCTATCATTTTA
TTGTGGTTTTGGTTGGGAGAGGAAATAAAATGGCCAATCCACTACCTCGAAAAAAAAAAAGCTTAAGGGCGAATT
C

SEQ ID NO. 83: **NC34-1-A,B,C**

GAATTCGCCCTTTCCGCTCTGGGGATATCAAAACTCTCTAGGTCCAGGTTCAAAATCTTCCACACATTCTCTGTG
TCTGCTTTTAGCCAGACACCATCACTATGTGGTAGCTTACCTCAAAGCTTCACTTAGTGTATCAACCCTC
AGAAGCACTTTCTGATCCCTTCAACCTGCACATCTGTCTTCTTATCTAAATTCCAGCCCAGCTCAATCCA
TACCTTCTGACCATGCTACGAAAAAAAAAAAGCTTAAGGGCGAATTC

SEQ ID NO. 84: **NG2-1-C,F,G**

GAATTCGCCCTTTGCCGAGCTGGGGAGTATAAAATGTTACCTCATTGTGGTTTCATTTTGTAAATTT
TTGATTATTAGTAAGTTGAACATGTTTTATTATGGGAATTCTTGTTTCTTTTTCTTGGCAATCCCTCTTC
ATGTCTTTTGCCTTTCTTCCTATTGAGTTGTGTTTTAATGATTAATTTTTAAAGTTTCTTTATATTTAAT
ATTTAATTGGTCGACATTTATTTAATCATTAAAGTGAAGAGAAACCAGATTTAGAGTAGA
AAACTTTTCTGAGGCCATTTCCGGAAATATGCTAAGCATGTGAATCTTTATTCTATTTGGAGAAAATAAA
GTTAAATACATATATAAAAAAAAAAAGCTTAAGGGC**GAAT**

SEQ ID NO. 85: **NG2-1-D**

GAATTCGCCCTTAAGCTTTTTTTTTTTACCAAACTATATTTACTGTTTACAGTAGTAAAGGACAACAAATAA
TGTACAAATTGGTGAATAAACACAACAGAGCCAACAAACATCCCACCCGAGCCCATACAGCAAACAGGAA
ATGAGAACATTTCAGCAAGATTTCAAGCAAGCAAGAGATGATGGGTCATTGTTCAGGTGACTGTAAAAAG
GCAGAGAATGGCCACCAAGCAGCAATGGAGCCTCAGGGAAGGACAATAGGCAGAACTATGAAAATGTTTA
ATTGGTATAGATCCCAAAATATTTCACAGAACTGAAATCACCAGACTAATGCATAAATTCAATACCTATT
TGGAAAGCAGCTCGGCA

SEQ ID NO. 86: NG2-2-I

GAATTCGCCCTTTGCCGAGCTGTTGTATATTGAGGTGTATTATTTACGTCTCTGGTCCAGTCTTTTCT
GGCAAATAACAGTAAAGATGGTTTAGCAGGTCACCTAGTTGGGTCAGAAGAGTCGATGATCACCAAGCAG
GAAAGGGAGGGAATAGAGGAATGTGTTCGGGTTAAGTGATGAAAATGGCAGTGGTGGCCGGGCGTGGTGG
CTCTCGCCTGTAATCTCAGCACTTTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCAAGAC
TAGCCTGGCCAACATCATGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCACAC
ACCTGTAGTCCCAGCTACTCGGGAGCCCAACGCACGAGAACCGCTTGTACCCAGGAGGTGGAGGTTGCAG
TGAGCCGAAGTTGCACCATTGCACTCCACCCTGGGCGACAGAGCAAGATTCTATCAAAAAAAAAAAGCTT

SEQ ID NO. 87: NG2-2-M

GAATTCGCCCTTTGCCGAGCTGGAAAAGACAATTGATGACCTGGAAGAGCGTCTCTACAGCCAACTTGAGCGAAA
C CGCCTGCTTTCTAATGAGCTGAAGCTAACGCTGCATGATCTGTGTGACTGATGGGCAGGGCTCAATGATG
CCCATTAAACTGAGCTTACTGCTCACACCACTGACCTGGACCCCAACAAAAAGCTGATTGTCTTTTTAAA
AGTTATTATTTTGCCCTGAGCAAATTGCATTTTAATTGGGGCAGTTAGAATGTTGATTTCCTAACAGCAT
TGTGAAGTTGACCATTGTGAAGTTTCTGTCCCTTTAGAAGAGATTATGGGTGAAGAAGGGAGGGGCCTGA
GAGATTATAGTGAGAAAACTTGCGAGAATTTTGTTTTCCACCCTTATTTGCTGCTCTTTCACTTGGGCAC
TGACTGTAGGATATGTTCCCTTGCATGGATGTTTTTAACAATAAAAGGACTGACTTGACAAAAAAAAAAA
GCTT

SEQ ID NO. 88: NG2-2-N

GAATTCGCCCTTAAGCTTTTTTTTTTTTGGAATCTACTGCGAGCACAGCAGGTCAGCAACAAGTTTATT
TTGCAGCTAGCAAGGTAACAGGGTAGGGCATGGTTACATGTTTAGGTCAACTTCCTTTGTCGTGGTTGAT
TGGTTTGTCTTTATGGCGGGGGGTGGGGTAGGGGAAAGCGAAGCAGAAGTAACATGGAGTGGGTGCAGCC
TCCCTGTAGAACCTGGTTACGAGAGCTTGGGGCAGTTCACCTGGTCTGTGACCGTCATTTTCTTGACATC
AATGTTATTAGAAGTCAGGATATTTTTTAGAGAGTCCACTGTTTCTGGAGGGAGATTAGGGTTTCTTGCC
AAGATCCAAGCAAAATCCACGTGAAAAAGTTGGATGATGCAGGTACAGGAATACACGAGGGCATAGTTCT
CATAGTCGGTGGCCAGGATCCAGTACGGTGCCGATGGCATAAACCAGGAAAACTTAACTTCCAGCTCGGC
A

SEQ ID NO. 89: NG9-1-F,H

GAATTCGCCCTTAAGCTTTTTTTTTTTTGGGCAGTGAGAGTGAGTAGTAGAATGTTTAGTGAGCCTAGGGTGTT
GTGAGTGTAAATTAGTGCGATGAGTAGGGGAAGGGAGCCTACTAGGGTGTAGAATAGGAAGTATGTGCCT
GCGTTCAGGCGTTCTGGCTGGTTGCCTCATCGGGTGATGATAGCCAAGGTGGGGATAAGTGTGGTTTCGA
AGAAGATATAAAATATGATTAGTTCTGTGGCTGTGAATGTTATAATTAAGGAGATTTGTAGGGAGATTAG
TATAGAGAGGTAGAGTTTTTTTCGTGATAGTGGTTCACTAGATAAGTGGCGTTACCCAAGGGCGAATTC

SEQ ID NO. 90: NG9-1-G

GAATTCGCCCTTGGGTAACGCCCAAGGAATAAGCAAGAACTGTGAGAGTCTGAGTAAGAAGCAGGGC
CAATCTCTGACCAGAGCCAAGGGCTCCTCTAATGATTGCACTTGTAAATTAGAAGGGAAATCTCTGTACT
ATAGGTATTGGTCATATCCTGAAGGAAGTCAAACATGCGCAGTCAACCTTTTTTAAACTCATGTCTTGAG
GAGAATATAGAGAATTGGTTACAACTGATTGGATACAACTGTATTCACAAGGTTAAGAATAATTTCAGTC
AAAATTAGAACAGGAAGTCTCTAGACTTTAGAAGTAATGTCATCCAAAAAAAAAAAGCTTAAGGGCGAAT
TC

SEQ ID NO. 91: NG10-1-A

GAATTCGCCCTTGTGATCGCAGAAGACAAGAAATAACCAACGTCAGAGCTTAACTGAAGAAGGAAATTGAG
ACTCAGAAAATTATCCAAAAGTTCAACAAATCCAGGGGTTGATTTTTTGAAAAATTTAATAAGATAGATA
GACCACTAGCTAGACTAATAGAGAAGAGAGACAATCCAAATAAACACCATTAGAAATGATGAAGGCAATG
TGACCACTGATCCCCACAGAAATACAAGTAACATGAGTAACTACTAAAAATGCTTCTATGCACACAAACTA
GAAAACCTTCTTAGAAGAGATAGGTAAATTTTTCTACACATACACCCTCCCAAGACTGAATCAAGAAGTA
ATTGAATGTTTTAACAGACCAATGATGAGCTCCAAAGTTGAATAGGAATAAATAAGCTACCAAAAAAAAA
AAGCTTAAGGGCGAATTC

SEQ ID NO. 92: NG10-1-B,D

GAATTCGCCCTTAAGCTTTTTTTTTTTTGGCAGATGGAGCTTGTTATAATTATGCCTCATAGGGATAGTACAAGGA
AGGGGTAGGCTATGTGTTTTGTCAGGGGGTTGAGAATGAGTGTGAGGCGTATTATACCATAGCCGCCTAGTTTAA
GAGTACTGCGGCAAGTACTATTGACCCAGCGATGGGGGCTTCGACATGGGCTTTAGGGAGTCATAAGTGGAGTCC
GTAAAGAGGTATCTTTACTATAAAGGCTATTGTGTAAGCTAGTCATATTAAGTTGTTGGCTCAGGAGTTTGATAG
TTCTTGGGCAGTGAGAGTGAGTAGTAGAATGTTTAGTGAGCCTAGGGTGTTGTGAGTGTAAATTAGTGCGATGAG
TAGGGGAAGGGAGCCTACTAGGGTGTAGAATAGGAAGTATGTGCCTGCGATCCAAGGGCGAATTC

SEQ ID NO. 93: NG24-1-P,Q

GAATTCGCCCTTTGCGCCCTTCCTGATCTATTTTCCATCTTTATCTCTCTTGTTCTGTGTTCCTGGGAGGCTGAC
AAATAAAGAACTGCATACTCTCTTTTCTGGCTTTATTCATGGTGGGTTTGGACAGTGATAGAAAAGTGAGGCTGA
AATATTTGCTCCTTCTCTTTCTCATATACTTGGCAAGGTCCAAGGGCGAATTC

SEQ ID NO. 94: NG25-1-M

GAATTC

101	GCCCTTTGCG	CCCTTCCCCC	AAATGTAGAG	TAAAAATCAT	ACTGAGGGAG
151	TTCAGGTTGT	TGCTCAGTGG	TTAACGAATC	TGATTAGCAT	CCATAGGATG
201	CAGGCTCGCG	CCCTGGCCTT	GCTCAGTGGG	TCAAGGATCT	GGCGTTGCCG
251	TGAGCTGTGG	TGTAGGTTGC	AGATGTGGCT	CAGATACTGC	ATTATTGTGG
301	CTGTGGTGTG	GGCTGGCAGC	TATAGCTCTG	ATTCAACCCC	CATCCTGGGA
351	ACCTCCATAT	GCCTTGGGTG	CCCTAAAAAA	AAAATCATAC	TTCAGAACCC
401	TCATGATCCT	GATGTTCCTT	CAAGAACAGT	CCTCTATGAG	TTCCTGCTGT
451	GGCACAGTGG	GTTAAGAATC	TGATTGCAGC	GACTCGGGTT	GCTGCTGCAG
501	AGGCAAGAGT	TTGATCCTCA	GCCTGGTATA	GTGAATTAAA	GGATCCAGCA
551	TTGCCACAGC	CATGGGGTAG	GTTACATCTG	TGGCTCAGAT	TCCGTCTCTG
601	GCTTGGGAAC	CTCCATATAC	TTGGCAAGGT	CCAAGGGCGA	ATTCCAGCAC
651	ACTGGCGGCC	GTTACTANTG	GATCCGANCT	CGGTACCAAG	CTTGATGCAT
701	AACTTGAGTA	TTCTATAGNG	GCACCTAAAT	AGCTTGGCGT	AA

SEQ ID NO. 95: OA2-1-A,C,L,N 2E4

GAATTCGCCCTTAAGCTTTTTTTTTTTACACTGTCTAAAGCATTTAATGGTCTTTCTTTAACACAGCCAACTCCC
CCGGGTTTGAAACAGTGTTAAATTCTCTCTTGCTTGTGGCAAAAGAAGCTGTCAAGTCCAACACTGAAAAATTGG
TACCATTTCCTGGCCAGTAAGCACAGAACAGAGGGGCTAAATATTTTATGGTTTTATTTATTTACTGTGTTCTCA
TGCTGTGTTTTTCTTTTCTCTGTCTCTCCCTCCTGCTCGTGTCTGCCCAGGGCTGATTGTTGTGACATTGGCCGT
ATGCTGGATGCCCAACCAGATTCGGAGGATCATGGCTGCGGCCAAACCCAAGCACGACTGGACGAGGTCCTACTT
CCGGGCGTACATGATCCTCCTCCCCTTCTCGGAGACGTTTTTCTACCTCAGCTCGGCAAGGGC**GAATTC**

SEQ ID NO. 96: OA2-1-M 2E3

GAATTCGCCCTTAAGCTTTTTTTTTTTTACAGAAGGTCAAACAAATGTATTATATGATTCAAATGGGACTATACAT
CTATTCATTTTTTAAGAGATAGGAGTGAAACAAATGAAAAAATCAACAAAGTACGTGCTTCTATAAATGAAGATA
ATTCCCAAGTTAAGCTACTATATAGTAAGAAATACCATATGCAAACTTCTAGACCACACAAAATTGGGGAAAAAT
TTTATCAAACTTATTAAAAAATAGCATTCATATCAATGTTGCATAAATACAGGAAAATATACAACCCAATAGAAA

TGTGACTAATGAGTATAAACACACTATGAATAGAATACAAACTCTAAACAGATTTAAGAGCACATATTCTCAACA
TCGTAAGTAGTTGGGGAAATGCAAATTAAATAAAACATTTTTCATCTAGCAGCTCGGCAAAGGGC**GAATTC**

SEQ ID NO. 97: OA3-4-1-G 2G7

GAATTCGCCCTTGACCGCTTGTTAAGAGGAACTGATCTCATATATTTGTATCAGAACTGTATTTTTATGTTATAT
TGTATAGTTTGCTCTCCTGCCCCTCTCCTTAAAACTGAATGGTGCCAATAATTTGATACTAATGACTACAAAAAA
AGGTAATGCCTCATTTACTAGTATTGTTGTAAAATGAGGAATGTATGTGAATATTCAGATAACCGAGGATTAACC
CTTTAAGTGCTGAATCTTTAAAATTTTAATATATTTTTTTTTGAGGGAAATCTTTCTAAAATGTATTACGCACTT
CCCTGCCTTAGTAAACAGAGTATACTGGAGAGTATTTAACCTTTTCTTGATGAGTCATGGTCATGATTATAAACA
TCAGCCCCTTTTAAAAAAAAAAGCTTAAGGGC**GAATTC**

SEQ ID NO. 98: OA10-1-B 2F1

GAATTCGCCCTTGTGATCGCAGGAATCAGGGGAAAGTGATTTTAAAGGTGGTTTCTCCAGCACATTTTAAGAAAA
GGGACCAAAAGTTATTTTAGCTTCCTCAATAGATTGCATGTTGCTTATTAGGATAATAAATTAATATTAAATGCA
ATATATGTCTTGTCTTTATTATGGCATCTATTTAGGAGTTGTTCAAATCACTGCAGTAGGGCTCTGCAAATAAAA
TAATGTAACCTATTATCATGGATCTAATGTACTGTAACTTTATCAGTGAAAGGTAAAATCTCAAATAACAAGTAC
AAACATTGAACAATTACCTATAAAGATTTGTAAAAGTAAAATTTTTCCAATAGATTTCATTCTTGTCATTTTGTA
AGACGACCCTGCAGTCCACCTGTTTGTAACTTTTTTAATAAAATAGACATCTGTAAAAAAAAAAAGCTTAAGGGC
GAATTC

SEQ ID NO. 99: OA10-1-E 2F2

GAATTCGCCCTTGTGATCGCAGGCTGTAAGTCTTTAGCATCTCAGGAAGTTACTAACTTCAAACTAAGTATATAG
GTAGAGTTTCTTACTAAATCTAGTGCTTCTTGAACCACAAGTAGAAAGCATTTAAAACATGAATGTTGTTTTGTG
TTTTTTGAAGTTTGTAAATAGAAGACTTGTTGATGATCCGATGGCAAGGTATTTTTCTCTTGGTATGTATTTTAG
TTATTTCCTCGTGATGCATAAGTGAAAAGAGTGAAGTTTCTCAGAATGAGCAACTGTCATCCATCTACCTGCTAT
TTTATTATTGCTGATTACAAAAGCAAATCAAGAGATGAGAACCCAGTTGCCTGCAAGTAAATATTTACTGCATTG
AGGGTCGGAGCATTTTCCCATTACCGGTTATCCATGGATCAAATAGTGTATCTCAGTGGTAATTCTAGAGGGCCA
TTAAAACCCTGATGGTGCTGGAAGAGATGGCAGTGCTGCATGTCAGAAATAGGTAAACTGTAATTAAGAAGTTAC
AGATGATTTGATTACGCTCTTGNGTATTTGGTCCTGTTATAATGTGAGCAGATTAAAAATCATGTAAGTGCTTAA
AAAAAAAAAGCTTAAGGGC**GAATTC**

SEQ ID NO. 100: OA10-1-H 2E12

GAATTCGCCCTTGTGATCGCAGTATTCCTTGTATGGAAGTCATCAGATATGCTGTGCAAGTCTTGCTTAATGTAT
CTAAGTATGAGAAAACTACTTCAGCAGTTTATGATGTAGAAAATTGTATAGATATACTATTGGAGCTTTTGCAGA
TATGCCGAGAAAAGCCTGGTAATAAAGTTGCAGACAAAGGCGGAAGCATTTTTACAAAAACTTGTTGTTTGTTGG
CTATTTTACTGAAGACAACAAATAGAGCCTCTGATGTACGAAGTAGGTCCAAAGTTGTTGACCGTATTTACAGTC
TCTACAAACTTACAGCTCATAAACATAAAATGAATACTGAAAGAATACTTTACAAGCAAAAGAA**GAATTC**TTCTA
TAAGCATTCCTTTTATCCCAGAAACACCTGTAAGGACCAGAATAGTTTCAAGACTTAAGCCAGATTGGGTTTTGA
GAAGAGATAACATGGAAGAAATCACAAATCCCCTGCAAGCTATTCAAATGGTGATGGATACGCTTGGCATTCCTT
ATTAGTAAATGTAAACATTTTCAGTATGTATAGTGNAAAGAAATATTAAAGCCAATCATGAGTACGTAAAAAAAA
AAGCTTAAGGGC**GAATTC**

SEQ ID NO. 101:1OA17-1-A,B,C

GAATTCG

101	CCCTTAAGCT	TTTTTTTTTT	AATGGTTGAA	GTAACTGAAG	ATATTTAATC
151	TAGAGAAGAT	TTGGGAAACA	CATGATAGCT	ATGGTTAAAT	ACTTAACAGG
201	GCAATCACAG	GGAAGATGAC	TAGATTTCCT	AACATCCATG	AGTGAAATTT
251	ATAGAAGTAT	ACTCTCTGAC	TTGATATAAA	GGAAGATTTT	AAAAAACATG
301	ACTGTTCAGG	AGTGTTCAAG	TAGGGTCAGA	TGACCAGTGA	TTGGGAATAC
351	TTCGTAAGCA	GGAGCAAGTA	AGATCTGAGC	CACTGTTCTA	TCGGTAGGGT
401	GTCTGTGGTA	TTCCTTGGTC	AAAGAAGTAC	TCTAAGCAAC	TTCAGTCTCA
451	CGAATTACTA	TCACCCTCGT	GGGCATACAT	GATGGTTACC	CTAAAGAGGA
501	AGTTTCAGAA	GGCAGTAATA	TTGGATCCTG	GAATAGTCAG	ACAGGAGCCT
551	TCATGCAGAT	ACCCTTTTCA	GTTCTCCATA	CACCCATTCA	CAAGCGGTCA
601	AGGGCGAATT	C			

SEQ ID NO. 102: OA17-2-A,B

GAATTCG CCCTTGACCG

101	CTTGTGAGGA	GGAAAGTAAT	GCTGGGAAAC	TTGATATGTG	TAAATAGAAA
151	ATATATAAGC	AAAGTTATCA	GCCAGTCTTG	ATGTTGCAGC	GGAAGTTGAG
201	AGTGCCGTGG	TATATCCTGT	TTTGTGCATT	AGCTTTTTCT	GGGGCATGAG
251	CATTCAGGCA	TTTTATGAAG	AACTTAGAAA	AAGTGAAAAA	TATTTTGAAG

```
301  TTTTATATCT  TTGATCATTA  GCTGGAAGGT  TTGTCCAGTA  GTAAGTTACT
351  TGTGAGGTTT  ATAAAATATT  AGGAACATTT  GGCAAGAAGA  GACAGGTTTT
401  GTGGGAATAA  TTTGTTACCT  GTTGACCCTC  ACTGTGGACA  TATTTGTGTG
451  TGTGTACCTG  TGTGTGTGTG  TGTGTGTGTG  TGTGTGTAAA  AAGGAGGGTT
501  TATAAAAAAA  AAAAGCTTAA  GGGCGAATTC
```

SEQ ID NO. 103: OA17-3-A,B,C

```
GA  ATTCGCCCTT  AAGCTTTTTT
101  TTTTTATACT  GCCTTGGAAT  ATTATCAGAT  GGATAAATTA  GCAGCACTTA
151  CTCCGCTGCA  TAAAACAAAA  TAAAATCATT  TGACCTAGGC  AAAAAACTTA
201  ATTTTGAAAT  ACTATGTGTT  TATTATCTCT  AACAGCTCTA  ATATTTCCTT
251  CTCAATAAAA  TTTAAATTTT  AAAATAAGCA  TTAAATAACT  AGCATTCAAT
301  GCCAACCACA  GTTCAGCATA  TTCATTTAAA  CCTTATTACA  TTCTGATACT
351  GGTATTGGTG  ACTTCACTTG  ATAATAAACT  GTTGCTCAGA  AAGGCCAAGG
401  ATTCTGTCCA  AAGTCCCAGA  GCTGGGAGCA  GCTGAATTAG  GAGTTGGCCT
451  GTCTGATTAT  GTAGCCGGGG  ATGTTTTCAG  TCACAAGCGG  TCAAGGGCGA
501  ATTC
```

SEQ ID NO. 104: OA17-5-B 2F4

GAATTCGCCCTTAAGCTTTTTTTTTTTAGCAGTAGAAATAAACCTAATTACTCATAAACCATATTTTGAAATGAG
AATAAATCAACAGCTTCATTTTGGAGCCTTTTAGAGTGCTAGAATATCTGGCCAAGGTAGACTGTGAAAGGTAGG
CTCTTCTTTAAACACGGTTATGGTTCAGCAGTTATTTGCAGGTCCTTGGGAAGGCACTGTGCTGAAGGAGAGCAA
AGAGTTTCTTTTTGTGCTTTTTTTTTTTTTGAGGNGGAAATATTTCTTGTGACAGTCAAGGCTCTTTTCTGATTT
TTTGTTGCTCATACAAGCGGTCAAGGGC**GAATTC**

SEQ ID NO. 105: OA17-5-E,F,G 2F9

GAATTCGCCCTTGACCGCTTGTGGCAACTTAGTGGAGTATGTTCCCTCTCAGGTAATATACAGAGAAGACAGGTT
AGAGGGTCTGTCTGTGAGTGTAT**GAATTC**CTTTTAGATTGGATGACTGATTTTTCTTACTTTAGTAAAGTTTTCA
AGTGCATGTGGACTGAAGGGCAGTAAGGAGGGCATAGAACAGCTATGG**GAATTC**CTAAA**GAATTC**ATCAGAGATG
AATGTAACGATTATGGAGTGAAGTATTTGAAATTTTGAAGTTAGCAGGGTTTTGTACTGTGCCAGTCTTTCATGA
TTTAAAAAAAAAAAAGCTTAAGGGC**GAATTC**

SEQ ID NO. 106: OA17-5-H,J 2F5

GAATTCGCCCTTGACCGCTTGTTGATGTCTACGGAAAGTGTGCTAGAATTTTAGTTAGGATTGTGTTGTGTCTAT
AGATCAAATTGTGAAGAAGTCACATCTTAGCAATACTTAGTTTTTGATTCCATGAACACAATATAATCTAATCAG
CACATAGATCCTGTACATATCTTACTGGATTTATACCCAAGTATGAACTGTTTTCTGGTTTTTTTTTTGTGTGTG
GGGGTGGGGGGACATGTGCTATTTGAAATGGTATATTCAATTGTTCATTGCTGGTATATGGAAATAAAATTGGCT
TCTGTAAAAAAAAAAAAGCTTAAGGGC**GAATTC**

SEQ ID NO. 107: OA19-1-B,D,G

```
G  AATTCGCCCT  TAAGCTTTTT  TTTTTTAAGA  TTGTTCTAAT
151  TCTGGTTGTA  AACTGCTATT  TTAAAAAACA  AAACAAACAG  AAAACATCAA
201  AAACACAAAA  AGATATTAAA  ACAGCAAGTC  TTTTGTACAT  CACTGTAGCA
251  TAAGCTGCTT  GAGGTTGTCA  TGCAGAATAG  TATCCTTCAC  GTCACGGAAA
301  ACAAGGCGGA  TGTTCTCCGT  GTTGATAGCA  GTGGTGAAGT  GGTGGTATAA
351  GGGCTTCTGT  TGCTGGTCCC  GACGTTTGAA  GGGC**GAATTC**
```

SEQ ID NO. 108: OA19-2-B,E

```
G  AATTCGCCCT  TAAGCTTTTT  TTTTTAAGAT
101  TGTTCTAATT  CTGGTTGTAA  ACTGCTATTT  TAAAAAACAA  AACAAACAGA
151  AAACATCAAA  AACACAAAAA  GATATTAAAA  CAGCAAGTCT  TTTGTACATC
201  ACTGTAGCAT  AAGCTGCTTG  AGGTTGTCAT  GCAGAATAGT  ATCCTTCACG
251  TCACGGAAAA  CAAGGCGGAT  GTTCTCCGTG  TTGATAGCAG  TGGTGAAGTG
301  GTGGTATAAG  GGCTTCTGTT  GCTGGTCCCG  ACGTTTGAAG  GGC**GAATTC**
```

SEQ ID NO. 109: OA19-3-D,E

```
GAATTCG  CCCTTAAGCT
101  TTTTTTTTTT  AAGATTGTTC  TAATTCTGGT  TGTAAACTGC  TATTTTAAAA
151  AACAAAACAA  ACAGAAAACA  TCAAAAACAC  AAAAAGATAT  TAAAACAGCA
201  AGTCTTTTGT  ACATCACTGT  AGCATAAGCT  GCTTGAGGTT  GTCATGCAGA
251  ATAGTATCCT  TCACGTCACG  GAAAACAAGG  CGGATGTTCT  CCGTGTTGAT
301  AGCAGTGGTG  AAGTGGTGGT  ATAAGGGCTT  CTGTTGCTGG  TCCCGACGTT
```

351 TGAAGGGCGA ATTC

SEQ ID NO. 110: OA19-5-D,F,G

GAATTCG CCCTTAAGCT

101	TTTTTTTTTT	AAGATTGTTC	TAATTCTGGT	TGTAAACTGC	TATTTTAAAA
151	AACAAAACAA	ACAGAAAACA	TCAAAAACAC	AAAAAGATAT	TAAAACAGCA
201	AGTCTTTTGT	ACATCACTGT	AGCATAAGCT	GCTTGAGGTT	GTCATGCAGA
251	ATAGTATCCT	TCACGTCACG	GAAAACAAGG	CGGATGTTCT	CCGTGTTGAT
301	AGCAGTGGTG	AAGTGGTGGT	ATAAGGGCTT	CTGTTGCTGG	TCCCGACGTT
351	TGAAGGGCGA	ATTC			

SEQ ID NO. 111: OA34-1-A 2G2

GAATTCGCCCTTAAGCTTTTTTTTTTTAAAGAAAAGATCTGCCCATCACCTATATTTTTATTTATCTCATGGGAT
TTTCGTATTTTCCTGGGAATGCAGGCACTCTGTTCTTATCATGGCTGAAATACGGTAGGCTTAATACTTCACAAT
TATATAGCACCTTTCACCCAAGGGCCTGTTGTTTGGTTTTGGTTTATGTGTGTTTTAATCAGCTTCCAGAATTGC
CATGCCTCACCTGTGAAGTGGGATAGGCAGGGTCCCCAAGAGGTGATCACTCCAGGTGGTGTCTAAGCCAGAGCG
GAAAGGGCGAATTC

SEQ ID NO. 112: OA34-1-B,C 2G4

GAATTCGCCCTTTCCGCTCTGGGACTATTACATTTAATTCTGCTCTTGATAGTCAAAGACCATGGACAACAACTG
TCATCTGAAGGACTCTTCTAGAAGCCAGAGACTGGTGTTTGATGGATGTTTTATACTAAATAAAACCCATCAGCA
TGGGGTTATGTAGAAAAGCAATTTATTCCATTTTAAGCACTTACACAGTTAGTCATGGAGAGTAACAGGCCTGCT
GGTGAAACAGGTCACCCAAAATGGAGATGGCATCAAACTAGTGGTCAAGGACTAACTCCTAAAAAAAAAAGCTTA
AGGGCGAATTC

SEQ ID NO. 113: OA34-1-F,H 2G6

GAATTCGCCCTTTGCTGTTGAAGAAGGCATTGTTTTGGGAGGGGGTTGTGCCCTCCTTCGATGCATTCCAGCCTT
GGACTCATTGACTCCAGCTAATGAAGATCAAAAAATTGGTATAGAAATTATTAAAAGAACACTCAAAATTCCAGC
AATGACCATTGCTAAGAATGCAGGTGTTGAAGGATCTTTGATAGTTGAGAAAATTATGCAAAGTTCCTCAGAAGT
TGGTTATGATGCTATGGCTGGAGATTTTGTGAATATGGTGGAAAAAGGAATCATTGACCCAACAAAGGTTGTGAG
AACTGCTTTATTGGATGCTGCTGAAGGGCGAATTC

SEQ ID NO. 114: OC13-1-A

GAATTCGCCCTTGACCGCTTGTGAATAATATTGTCTCTATAGGTGTGCAAGCATTTCCTGGAAGCTATTGAAAAC
AACAAGTATGGCTGGTTTTGGGTATGCCCTGGAGGGGGTGATATTTGCATGTATCGTCATGCACTTCCTCCTGGA
TTTGTGTTGAAAAAAAAAAAGCTTAAGGGCGAATTC

SEQ ID NO. 115: OC13-1-C,D

GAATTCG CCCTTCAGCA

101	CCCACTGAAA	AACAAGTTGA	GTAGAGAGTG	TAGAGTGCAG	AAATGTGGCT
151	TTTGCCCCAC	TTTGCATCTC	CAAAATTACA	ACGGTTGGCC	GATCCCATTT
201	GAGGACAATG	CTTAGTTATA	AGTCTCCGAG	TTGGAAAAGG	AAGAAAGCCA
251	GAGCTGTCTA	GTTTCATTCA	TTCTTTCAGT	AAATATTTAT	TGAGTACCTA
301	CTGTGTGCTA	GGCATTGACC	TGGGAACTAG	AGATACTTCA	CAGAATAACA
351	GGGAAAGTTC	CCTGTGCTCA	TGGAGCTTAC	ATTCTACAGG	GAGAAAGAGA
401	TAGCCAATAC	ATAGGAATAA	ATATATACAA	GGTATCATGT	AGTGATAATT
451	GCTGTGGAAA	AAAAAAAAGC	TTAAGGGCGA	ATTC	

SEQ ID NO. 116: OC17-2-B,C

GAATTCGCC CTTGACCGCT

101	TGTTAAATAC	CTTTTGCTAG	CCTCTTCATA	TGCTGTTGCA	TATGACTCTC
151	ATCACAACTC	AGTGAGATGG	AAAGACAAAT	CCTATTTGTA	CAAATGAGAA
201	AACTGAACTC	TTTAGAGTAA	CTAGCTCAGT	ATTGGCCAGC	TGGTAAATGG
251	CAGTGTTGGG	ATTAAAATCC	AGTTCTTATC	TACTCTCCCT	TTATTCAGAA
301	GCATTTATTG	GATGTTGATC	TTTGTTTCAG	GTTTTGATTT	TGTTACTTTT
351	TTATACTGTG	TATATTTTCC	TCAGTCTACC	CTTCTGCTCT	AGATTGTCTG
401	GACTCAGGAG	ATTGTGGCAG	TTACTGGATA	GTTATTTTTA	AGATAATGAT
451	TGCTTTTCTC	TGTTTATATA	AGTCATGTGT	ACTTATTGTA	GAAAGTTTGT
501	AAGATGCAAA	AAGTATAAAA	ATTAAAGTTA	TGCACTACTA	ACATTTCAAT

551 ATATTTTCTC CCAGATTTTC AATAAAGACT TTCAGGCAGT GAAAAAAAAA
601 AAGCTTAAGG GCGAATTC

SEQ ID NO. 117: OC17-3-A,B,C

GAATTCGCC CTTGACCGCT TGTCACTGAA TTGGTTTGCA
151 CACACTAACA TTTTACTCTA AAACAACTAA GTTGCATTGG AATCTGATGG
201 AATATATTGA AACATATCCG TGACCTTTGA ATTGTAAGTA ATAAGTTGTG
251 GAAAGTATAC TTAACTTGAC AGCATTAAAA ACAAATTAAT TTTGGTCTTA
301 TCTTAAGATT TGACTGCCTA TATAAGGTAG TGACTGACCT ATGAAAGCTC
351 TTTTATGTTG AAAGCAAGTG AAAAAAAACT AAAGCCTTAT TGGTTTGAGG
401 TTAGAACGGT TATTTGAAAA GTGGATTTGA AAAGAACTGA AGCTGAATTA
451 TTCTAAAAAC AAAGGAATGA AGCTTTATGA CAGGGCACGT GAAATGTTTA
501 TAGTGAAAAG GGAGAAATAA GTAACAATTG AAAAAAACTT CTA**GAATTC**C
551 ATTTAGTAAC AAAGAGGTTT TTGATGAAAA TTGTTTGGGA AAAAAAAAAA
601 GCTTAAGGGC **GAATTC**

SEQ ID NO. 118: OC17-7-A,E,J

GA ATTCGCCCTT GACCGCTTGT
101 GAATAATATT GTCTCTATAG GTGTGCAAGC ATTTCCTGGA AGCTATTGAA
151 AACAACAAGT ATGGCTGGTT TTGGGTATGC CCTGGAGGGG GTGATATTTG
201 CATGTATCGT CATGCACTTC CTCCTGGATT TGTGTTGAAA AAAGATAAAA
251 AGAAAGAAGA GAAAGAAGAT GAAATTTCAT TAGAAGATCT AATTGAGAGA
301 GAGCGTTCTG CCCTAGGTCC AAATGTTACC AAAATCACTC TAGAATCTTT
351 TCTTGCCTGG AAGAAAAGGA AAAGACAAGA AAAGATTGAT AAACTTGAAC
401 AAGATATGGA AAAAAAAAAA GCTTAAGGGC **GAATTC**

SEQ ID NO. 119: OC17-6-A,B,C

GAATTCG CCCTTGACCG
101 CTTGTGTGGA GAAGGGGAAT AGAGGTGAAT TTAGGCTAAC CAGTTAGCTG
151 AATGGTGGAC TCAGCCTTTG GCAGGAAAGA TTTAAGAGAA TTGGTTATTG
201 GAAGTAGGGG TGATGAGGGC TTGTGACTGA TTGCATGTGG TGAGCAAAAG
251 GGAAGAGGTG CCATGGGTGA TTCCCAGATT TTTGTTTGGC AGGGGTAACT
301 ACATTGGATG GATATGTCAA GATGGAGAAG GAGCAGATGA GGTAGGCATT
351 CATTAATTTT TATTTAAATA TTTTCCTTTG TTGGGTATAC CTGGAGTGTC
401 CTTTGTAGAG CCCCCGATTA GGCTCTGTCA GTGTGATAAA ACAAATAGTT
451 TTAACTAGAA AAAAAAAAAG CTTAAGGGC**G** **AATTC**

SEQ ID NO. 120: OC19-4-A,C,G

GAATT
101 CGCCCTTAAG CTTTTTTTTT TTCGTACTAC ACGACACGTA CTACGTTGTA
151 GCCCACTTCC ACTATGTCCT ATCAATAGAA GCTGTATTTG CCATCATAGG
201 AGGCTTCATT CACTGATTTC CCCTATTCTC AGGCTACACC CTAGACCAAA
251 CCTACGCCAA AATCCATTTC ACTATCATAT TCATCGGCGT AAATCTAACT
301 TTCTTCCCAC AACACTTTCT CGGCCTATCC GGAATGCCCC GACGTTTGAA
351 GGGC**GAATTC**

SEQ ID NO. 121: OC19-5-E

GAATTC
101 GCCCTTAAGC TTTTTTTTTT TCGTACTACA CGACACGTAC TACGTTGTAG
151 CCCACTTCCA CTATGTCCTA TCAATAGGAG CTGTATTTGC CATCATAGGA
201 GGCTTCATTC ACTGATTTCC CCTATTCTCA GGCTACACCC TAGACCAAAC
251 CTACGCCAAA ATCCATTTCA CTATCATATT CATCGGCGTA AATCTAACTT
301 TCTTCCCACA ACACTTTCTC GGCCTATCCG GAATGCCCCG ACGTTTGAAG
351 GGC**GAATTC**

SEQ ID NO. 122: OC19-5-F

GAATT
101 CGCCCTTAAN CTTTTGTTTT TTAANANTGT NCTANNNCTG NTTGTAAACN
151 GCTATTTTAN AAAANANANC ATNCAGANNA CATGAANANC NCANAANNAT
201 ATNAAAACAT CAAGNNCTTT TGTACATCAC TGTAGCATAA GCTGNTNGAG
251 GTTGTNANGC AAAATACTAT NCTTCANGTG ACGGAAAACA AGGNGGATGT
301 TNTCCGTGTT GANAGCAGNG GTGAANNGGT GNNATANGGG CTNATGTTGN

351 TGGGCCCAAC NNTNGAAGGG CGAATTC

SEQ ID NO. 123: OC22-1-A,B 2F11

GAATTCGCCCTTTGATCCCTGGATTATGCAAAGAAAAATGAACCCAAACATAGACTTGCAAGACATGGCCTGTAT
GAGAAGAAAAAGACCTCAAGAAAGCAACGAAAGGAACGCAAGAACAGAATGAAGAAAGTCAGGGGGACTGCAAAG
GCCAATGTTGGTGCTGGCAAAAAGTGAGCTGGAGATTGGATCACAGCCGAAGGAGTAAAGGTGCTGCAATGATGT
TAGCTGTGGCCACTGTGGATTTTTCGCAAGAACATTAATAAACTAAAAACTTCATGTGAAAAAAAAAAAGCTTAA
GGGCGAATTC

SEQ ID NO. 124: OC22-1-C 2F12

GAATTCGCCCTTTGATCCCTGGGACACATTCTCAAAAATAGTATTCCTTGGGCTTTATAGGAAGTCTGATGAGAG
ACAATGTGGCTTTATTAGAGTGGAGAAGGTGCAGATGAAGAGATAACAGGCTCCAGGCATGTTTTGGAGGCAATA
GGCTAGATTTTAGGGGAGAAGTAAACTAAGGAATTAAGATAGTTTTCAGGTTTTAGCTTTGAACAATTGGGTGGT
TGGTGGACACCGTTACTAACACTGGGAAACCTGAAAAAGAAAGATATTTGGGGAAAAAAAAAAAGCTTAAGGGCG
AATTC

SEQ ID NO. 125: OC22-1-D

GAATTCGGCTTAAGCTTTTTTTTTTTCAGGTAGAATTTTTTCTACAAAAATGGTGATTTATTTAACATACAATGG
TAGATATATTGGGGATCATCAAATTTTTAAAAATTTTTGTTGGGCTTACTTTACATATTACAATACTAACGTAAG
CTTATAAACTCTTCCATTTCAGTGAAGGAAAGGGATCCAATTAACACCTGCTAGCAGCTCGGCAAAGCCGAATTC

SEQ ID NO. 126: YA2-3-D,F

GAATTCGCC CTTAAGCTTT TTTTTTTTAT TGTAAATACT

151	CTTTATTGTA	AATATTCTAT	CCTAAATTCC	ATATAGCCAA	TTAATTCTTA
201	CAGAATCTTT	TGTTAATTTT	TGTGTGTATA	AATTTTACAG	AGATAAAGGG
251	TATGTTTGTT	GCACACAACT	TACAAATAAT	AATAAACTCT	TTATTGTAAA
301	TATTCTTTAT	TGTAAATTCT	TTATCCTAAA	TTCCATATAG	ACAATTGATT
351	CTTACAGAAT	ATTTTGTTAA	TATTTTTTTT	TTTTGCCAAA	CCTTGTATCC
401	AAAATCAACT	ATCACCGCGA	TTTGGCCATG	ATTTGACAAA	ATTAGACTAC
451	AAATAAAAAT	CCCTGATAAC	TGTACAGCTC	GGCAAAGGGC	GAATTC

SEQ ID NO. 127: YA2-3-E

GAATTCGC CCTTGCCGAG

101	CTGCTTTCCA	AATAGGTATT	GAATTTATGC	ATTAGTCTGG	TGATTTCAGT
151	TCTGTGAAAT	ATTTTGGGAT	CTATACCAAT	TAAACATTTT	CATAGTTCTG
201	CCTATTGTCC	TTCCCTGAGG	CTCCATTGCT	GCTTGGTGGC	CATTCTCTGC
251	CTTTTTACAG	TCACCTGAAC	AATGACCCAT	CATCTCTTGC	TTGCTTGAAA
301	TCTTGCTGAA	ATGTTCTCAT	TTCCTGTTTG	CTGTATGGGC	TCGGGTGGGA
351	TGTTTGTTGG	CTCTGTTGTG	TTTATTCACC	AATTTGTACA	TTATTTGTTG
401	TCCTTTACTA	CTGTAAACAG	TAAATATAGT	TTGGTAAAAA	AAAAAAGCTT
451	AAGGGCGAAT	TC			

SEQ ID NO. 128: YA4-2-B,C,D

AATCGG GCTGCATAAA TACACATTAT CTAATGTATT ATAATATTCA TAACAATCCT
CTGTGTTATC TATAGCCCAT TTCACAGGTA AGAAAAAAGA CTAAAAGACA
TTTAAGTGAC TTGATTAATG ACAAAAATAG GCAGTCAACC TGAACATCAA

251	CCCAAATATT	CTTATTACAT	CAATACAGCC	TCTCAGCAAC	CAATTAACTC
301	CAGAGAATGA	AAGAAGGGGA	AGAAGGAATA	AAGATTCTGA	GTGAGGGAAA
351	TAAATATGTA	TAAGACGTGG	CAGGTTTAAC	TTTAAGGAGG	TGGTATATTA
401	CTTTTGTCTA	AAAAAAAAAA	GCTTAAGGGC	GAATTC	

SEQ ID NO. 129: YA9-2-A,B,C

GAATTCGC

101	CCTTGGGTAA	CGCCATACTT	CATAAGTGGT	AAAGAAAGGT	ATAAAATTTG
151	GAAACATTTT	GTTGGGCATA	GTAGTGATTG	GGTGAAAAGG	ATAAATTATA
201	TCAAAATGAG	AATGTGCTGT	AATTGGAAGT	AGGGAGCTAA	AGGATGTTTC
251	TTTCAGTTTA	GTAGAACTGG	AACGTTTTAC	TATTAAACAT	GGCTTTTATA
301	AATGCATGGT	CCAATAATTT	TATTCACTGT	TAGTATTTAA	TTCACTGTCA
351	GCTTATTAAT	GTTTTCTGTA	CCCATTAATG	AATTTTAAAT	TACAAAAAAT
401	TGTCTAGCAG	CTACAGTTTA	AAAATGAAAC	TAGACATTAA	AATAAATTTG
451	ATAATTTTTT	ATAAAAAAAA	AAAGCTTAAG	GGCGAATTC	

SEQ ID NO. 130: YA9-3-C,D
GAATTCG

```
101  CCCTTAAGCT  TTTTTTTTTT  AATTTCACAA  AAGTTTTCAC  AAGGACAACG
151  TTATAGAAGA  AAACCCCCAG  CAGTGGCTAG  GTCATGCAGA  ACCATTAATT
201  GTCATACCTT  GGCCCATTCT  ATTCATCCTT  GTTGCACTTT  AGAGAGAGAA
251  GTAAGCTATG  TGAGTTTTAC  AATGCTTTTA  AACTGTCATA  TTTCCTGTTG
301  AGCACTTTAA  CTGGCACATT  CTTATAGTTA  TAAATGTTCT  GAGGGCGTTA
351  CCCAAGGGCG  AATTC
```

SEQ ID NO. 131: YA11-3-D,E
GAATT

```
101  CGCCCTTAAG  CTTTTTTTTT  TTAGGCCCAT  TTGAGTATTT  TGTTTTCAAT
151  TAGGGAGATA  GTTGGTATTA  GGATTAGGAT  TGTTGTGAAG  TATAGTACGG
201  ATGCTACTTG  TCCAATGATG  GTAAAAGGGT  AGCTTACTGG  TTGTCCTCCG
251  ATTCAGGTTA  GAATGAGGAG  GTCTGCGGCT  AGGAGTCAAT  AAAGTGATTG
301  GCTTAGTGGG  CGAAATATTA  TGCTTTGTTG  TTTGGATATA  TGGAGGACGG
351  CGATTGAAGG  GCGAATTC
```

SEQ ID NO. 132: YA20-4-C,H
GAATT

```
101  CGCCCTTAAG  CTTTTTTTTT  TTATAATGAT  AATTTTTATA  CTTTTATTAC
151  TTAGAAAATA  ATTTATATTT  TTCCATCATT  TAAACAAAGA  GTAGGCCTGA
201  GTCTCATGCC  TTTTGCACAG  CTTTTACCTT  CAAAGAAAGT  TATCTGGGTA
251  AGATAGGCAG  GAAATATGGG  GAAACTGCAA  ATTAACAGTC  TACATACATC
301  TAATATGAAC  AGTCTGTAAG  ATATTCCTTT  TCTTTCGTTT  TACTGGGATC
351  GCAACAAGGG  CGAATTC
```

SEQ ID NO. 133: YC1-5-F,H,I
GAATTCG CCCTTCAGGC

```
101  CCTTCGATGT  ATGCCATTTA  GTGAAAGTGC  TAAGTCTTAA  GTTTCCTACC
151  ACTTTGGTTT  CATATTTTTG  GACTTAACAA  AGTTGTGAAT  AGCACAGTCG
201  AGGAAAATTG  ATACCTGCAG  TAACCCATAG  GAAATAAACT  GTAGAGTTCC
251  ATATTCTGGT  ATTGTGATTA  TATTGTTTTA  TATTAAAAAG  GAAAAGAAAA
301  GAATTTTTTT  TAATTTTATT  TTTCCCCGTC  TTGCAAAGTA  TAGTGACCCC
351  TGTTTCCATT  AAATTTGAAT  AAAGACTATT  TTTGCTTGAA  AAAAAAAAAG
401  CTTAAGGGCG  AATTC
```

SEQ ID NO. 134: YC2-3-G,I,P,R,T 2E9

```
GAATTCGCCCTTTGCCGAGCTGTGGGGATCTGGCACTGTGGTTCCTGCATGAAGACAGTGGCTGGCGGTGCCTGG
ACGTACAATACCACTTCCGCTGTCACGGTAAAGTCCGCCATCAGAAGACTGAAGAAGTTGAAAGACCAGTAGACG
CTCCTCTACTCTTTGAGACATCACTGGCCTATAATAAATGGGTTAATTTATGTAACAAAATTGCCTTGGCTTGTT
AACTTTATTAGACATTCTGATGTTTGCATTGTGTAAATACTGTTGTATTGGAAAAGCATGCCGAGCTGGAAAAAA
AAAAAGCTTAAGGGCGAATTC
```

SEQ ID NO. 135: YC4-2-B,C,D
GAATTCG

```
101  CCCTTAAGCT  TTTTTTTTTT  CACGGAGGAT  GGTGGTCAAG  GGACCCCTAT
151  CTGAGGGGGG  TCATCCATGG  GGACGAGAAG  GGATTTGACT  GTAATGTGCT
201  ATGTACGGTA  AATGGCTTTA  TGTACTATGT  ACTGTTAAAG  ATGGGTAGGT
251  TTGTTGGTAT  CCTAGTGGGT  GAGGGGTGGC  TTTGGAGTTG  CAGTTGATGT
301  GTGATAGTTG  AGGGTTGATT  GCTGTACTTG  CTTGTAAGCA  TGGGGAGGGG
351  GTTTTGATGT  GGATTGGGTT  TTTATGTACT  ACAGGTGGTC  AAGTATTTAT
401  GGTACCGTGC  AATATTCATG  GTGGCTGACT  AAGGGCGAAT  TC
```

SEQ ID NO. 136: YC13-1-G,I
..GAATTCG CCCTTAAGCT

```
101  TTTTTTTTTT  CGGTTAGGGT  ACCGCGGCCG  TTAAACATGT  GTCACTGGGC
151  AGGCGGTGCC  TCTAATACTG  GTGATGCTAG  AGGTGATGTT  TTTGGTAAAC
201  AGGCGGGGTA  AGATTTGCCG  AGTTCCTTTT  ACTTTTTTTA  ACCTTTCCTT
251  ATGAGCATGC  CTGTGTTGGG  TTGACAGTGA  GGGTAATAAT  GACTTGTTGG
301  TTGATTGTAG  ATATTGGGCT  GTTAATTGTC  AGTTCAGTGT  TTTAATCTGA
351  CGCAGGCTTA  TGCGGAGGAG  AATGTTTTCA  TGTTACTTAT  ACTAACATTA
```

401 GTTCTTCTAT AGGGTGATAG ATTGGTCCAA TTGGGTGTGA GGAGTTCAGT
451 TATATGTTTG GGATTTTTTA GGTAGTGGGT GCTGAAGGGC GAATTC

SEQ ID NO. 137: YC13-1-H

GAATTCG CCCTTCAGCA
101 CCCACTGAAA AACAAGTTGA GTAGAGAGTG TAGAGTGCAG AAATGTGGCT
151 TTTGCCCCAC TTTGCATCTC CAAAATTACA ACGGTTGGCC GATCCCATTT
201 GAGGACAATG CTTAGTTATA AGTCTCCGAG TTGGAAAAGG AAGAAAGCCA
251 GAGCTGTCTA GTTTCATTCA TTCTTTCAGT AAATATTTAT TGAGTACCTA
301 CTGTGTGCTA GGCATTGACC TGGGAACTAG AACTAGAGAT ACTTCACAGA
351 ATAACAGGGA AAGTTCCCTG TGCTCATGGA GCTTACATTC TACAGGGAGA
401 AAGAGATAGC CAATACATAG GAATAAATAT ATACAAGGTA TCATGTAGTG
451 ATAATTGCTG TGGAAAAAAA AAAAGCTTAA GGGCGAATTC

SEQ ID NO. 138: YG1-1-J,K

GAATTCGCCCTTGGACCTTGCAAGTATATGTTCAAGTAGATGGCTTGCTGGTTACTGGACAAAATCCAGCCTCTT
CTGCTGCAACAGCTACAGCATTACTTCAATTGTTGAAATAAAGAGTAGCGAGTGCCTGTTGTGGTGCTTAGCAGT
GAATAAAACTTCTGACTCTGAGCAACTGGCGCAATAAGCAAAATAACCCTCATTTAATGGTGAAGGGCCTGAAGG
GC**GAATTC**

SEQ ID NO. 139: YG2-1-A,H

GA ATTCGCCCTT AAGCTTTTTT
101 TTTTTGGACA GGAAGTAGAA TTTATTGGTG AGTATTAAGA GGGGGGCAGC
151 ACATTGGAAG CCCTCATGAG TGCAGGGCCC GCCACTTGTC CAGAGGGCCA
201 CGATTGGGGA TGTACTTGAC CCCACAGCCA TCTGGGATGA GCCGCTTTTC
251 AGCCACCATG TCTTCAAATT CATCAGCATT GAACTTGGTG AAGCCCCACT
301 TCTTTGAGAT GTGGATCTTC TGGCGGCCAG GAAACTTGAA CTTGGCCCTG
351 CGCAGGGCCT CAATCACATG CTCCTTGTTC TGCAGCTCGG CAAAGGGCGA
401 ATTC

SEQ ID NO. 140: YG2-1-G

G
101 AATTCGCCCT TAAGCTTTTT TTTTTTGGCA GTGGTATAAC TATATTTATT
151 GTGCCTGAGA GGCAAGGTGA GGGAAAAATC TCAACAGAAG CAAGTTTGGG
201 GAAAATCTGG AGTCCCCAGT AAAAAGCAGG AAGGTCTCTG CTGTACTCAT
251 CACAGAATGG GAGAGAGGGC TCTCAATAGA TCATTCCCTT TGTTTCTCCC
301 CTGGGCTTCT TGAGCTTCTC GAAGTTCTTC AGGATGATGT CATATAACAC
351 AGCATAAGCA TTGCGGATCT CCATGACCAT CAGCCGGATG TCCCGGTACT
401 CTGCCTCATC CAGCTCGGCA AGGGCGAATT C

SEQ ID NO. 141: YG2-2-Q,R

GAATT
101 CGCCCTTTGC CGAGCTGCAG AACAAGGAGC ATGTGATTGA GGCCCTGCGC
151 AGGGCCAAGT TCAAGTTTCC TGGCCGCCAG AAGATCCACA TCTCAAAGAA
201 GTGGGGCTTC ACCAAGTTCA ATGCTGATGA ATTTGAAGAC ATGGTGGCTG
251 AAAAGCGGCT CATCCCAGAT GGCTGTGGGG TCAAGTACAT CCCCAGTCGT
301 GGCCCTCTGG ACAAGTGGCG GGCCCTGCAC TCATGAGGGC TTCCAATGTG
351 CTGCCCCCCT CTTAATACTC ACCAATAAAT TCTACTTCCT GTCCAAAAAA
401 AAAAAGCTTA AGGGCGAATT C

It should be noted that the sequence "GAATTC" at the 5' or 3' ends of SEQ ID NOS. 1-32 may represent a restriction enzyme site used in characterizing the sequences and does not necessarily constitute part of the differentially expressed sequence.

Example 3

Age-Related Differential Gene Expression in Glioblastoma

1. Patient Characteristics

The 211 patients were diagnosed with glioblastoma multiforme (GBM) at the Chicago Institute for Neurosurgery and Neuroresearch between October 1, 1987 and December 30, 1994 and consisted of 94 females and 117 males. 180 patients had lesions confined to one cerebral hemisphere and 30 had lesions that were more extensive or were multifocal. All tumors were classified by the same neuropathologist essentially according to a four-tiered grading system typified by that of the WHO Classification scheme. Patients with high grade oligodendrogliomas, and mixed cell gliomas were excluded from the study. Survival was measured from the date of first surgery at CINN to the patient's death or the date of the last clinic visit, and updated to March, 1996. The strength of association between the survival times of different patient groups was determined using the modified Wilcoxon test.

2. Tissue Materials

For the Differential Display analysis, 3 GBMs excised from older (>60 yr.) patients, 3 GBMs excised from younger patients (<45 yr.) and 3 sections of normal gray matter were used. Their individual characteristics are listed below:

Normal	"Young" GBM (<45yr.; "Y")	"Old" GBM (>60 yr.; "O")
UMB 242 (46 yr., female)	CINN 319 (37 yr., male)	CINN 407 (64 yr., female)
UMB 418 (53 yr., male)	CINN 361 (40 yr., male)	CINN 422 (64 yr., male)
UMB 389 (71 yr., male)	CINN 504 (43 yr., female)	CINN 419 (72 yr., female)

Normal human brain tissue was obtained from the Brain and Tissue Bank for Developmental Disorders at the University of Maryland, Baltimore, Maryland. Brain tumor tissue was obtained from the tumor bank maintained by CINN.

3. Differential Display

Total RNA from 3 individual specimens per patient group was extracted by guanidinium thiocyancate followed by cesium chloride sedimentation (Chirgwin, et al.) and treated with DNase I. Reverse transcription was performed utilizing single base anchored primers: (T11M, 5' TTTTTTTTTTTM 3', where M denotes A, C or G). Differential display was performed

essentially as described (Liang, et al., 1992). For each of the three anchored primers in each sample, 28 arbitrary upstream primers were utilized in the PCR amplification to produce a total of 84 unique primer pairs in the analysis. The resultant amplicons were electrophoresed on 6% sequencing gels. Differentially expressed amplicons were excised, reamplified and purified. They were subsequently subcloned into the TA cloning site of the pCR2.1 vector (Invitrogen, Carlsbad, CA) and insert-containing vectors from multiple positive transformants sequenced using an ABI 377 automated fluoresence-based sequencer. All NCBI maintained nucleotide databases (National Center for Biotechnology Information, Bethesda, MD) were searched for homologies using the BLAST program (located at http//www.ncbi.nlm.nih.gov/BLAST/index.html).

4. **Northern Blot**

25 mg of total RNA isolated as above was electrophoresed through 1.2% formaldehyde-agarose gels and transferred to nylon membranes by capillary blotting. The membrane was hybridized to a uniformly (^{32}P)-labeled hsp60 cDNA amplicon identified by differential display. This probe is homologous to the 3' end of the hsp60 protein coding region. Filters were hybridized for 90 minutes at 68°C using Express-Hyb (Clontech, Palo Alto, CA). Filters were washed in 0.1X SSPE/0.1% SDS at 50°C and analyzed by autoradiography for appropriate times. CDNA probes for genes representative of the other major stress protein families, hsp27, hsp70, hsc72, hsp89α, hsp89β, and GRP78 were generously provided by Dr. Richard I. Morimoto.

5. **Sequences identified as being over-expressed in "Old" tumors**

Using the DDRT-PCR methodology described above, the following sequences were identified as being differentially expressed (ie, overexpressed in "old" tumors as compared to "young" tumors) in tumor cells taken from patients older than 60 yrs, of age.

SEQ ID NO. 142: **OA 3-1-B**

AGTCAGCCACCATGAACAAAGTGGATCTTGTCTTCTTACATCTATGAAAATAGAGCTTTGAA
TGGTAAGGAGATATGTTTTCTTGGTAACCAATGCAAGATTGATGGGTGGAAACATGATTCAA
ACTTACACAATTTTTCTTGCTATTTTTCAAATATGAATCTTACTATATATTCTCGGTGAACA
TCAGGAGACTATTAAAGAGGTCTGCTGTTAAATGTAAAAAAAAAAAGCTT

SEQ ID NO. 143: **OA 11-4-1**

AAGCTTTTTTTTTTTAGAAATCAGGNGKTTTTTTATTTAATACATTCTAATCAAATAGTAAC

AGCAGTAAATAAACACTTTGAAAAACAGGCAGGTATCCCCCTGTATCTGGAAGAAAATTAAG
TCAAAGTATTCTACACAGTAGAAGGGAGACAACTGTTTATGTCCATGGTTAGACAATTCAAG
GACAACTTGGATATTTCTAAAGCCATTTCCAAAAAATCAATGGCAACAGGTTGGGACACAGC
TATTTCAAAGGGTAGAATGCCTATACCTACATTGGTTTTTATTAACGGCGATTGAAGCCGAA
TTC

SEQ ID NO. 144: **OA 11-5-C**

AAGCTTTTTTTTTTTAGCGACAGTTGTATTTATTTTTTTAAGTTACAATAAAATGCTCTCAA
GTCCTTTGAATGTTCCAACAAATTCAAAACTTCATTTTCTGAATGTTTTACATAAATGCGAA
CTACCTGTTCGCATTGGNAACCTGCTGCTGTATTTCATGTCTTAACGGCGATTG

SEQ ID NO. 145: **OA 11-6-A**

AAGCTTTTTTTTTTTACAAATGGAAGGTTTCTGACAAACTTAAGTGGAGCAAGTACAAGTCT
ATCAGTGCAATTTTTCCAATAGCATATGCTTACTTCCTATGTGTCATGTTTTGGTAATTTTC
ACAAAATTTAAACTTTATTACTATTATACCTGTTACGGCGATTG

SEQ ID NO. 146: **OC 11-1C**

CAATCGCCGTCATGGAGTGCAATAATGAGTGAAAAAAGTTTGATATTATCTATGTAATGAGT
TGATAACGACCTATTTTTTTTTAAAGAAGTCTTGCCTTTAATAAAAACCTCAACTATAACAT
GTGGCACTTGATGTACATTCGCGTTCCATCTTCGTAAAAAGCCTGTGGAATAGGTAGGTATT
ATCTTTTATAGATGTGGAAATGTAGGCTTCGTTATTTTAATAGCTTGTCGAAGCTTTACACA
GGTAGTAAGAGGCAGATTTGAACCTAGGCATTCTGATTGCAAGTAATTTCCCTTTCATTATG
CCACAGTGTGTTTATTATATACACTGAGTGTAGCTAATCGCCACTGGAGACGCCTTTGGAAA
AAAAAAAAGCTT

SEQ ID NO. 147: **OC 11-4-C**

AAGCTTTTTTTTTTTCGAAGGAAAATTTGTATTATTTSAATTATTTTTATGKACAGAAAACT
CAACAGTGTACATTTAACCCAGTTTAGKGGCAAGTTCTTTAGCCTTTGCCTTTTCGAGCTTG
GCGATACGAGCCACAGACTTAGGACCCAGGACACTGCCACCCCAGTGACGGCGATTG

SEQ ID NO. 148: **OC 12-3-3**

GCTGATAGTGACTATGGCAGTTCGAAAAAAAAAAAGCTTAATATAGCAAGGACTAACCCCTA
TACCTTACTACCAGACAACCTTAGCCAAACCATTTACCCAAATAAAGTATCGGCGATAGAAA
TTGAAACCTGGCGCAATAGATATAGTACCGCAAGGGAAAGATGAAAAATTATAACCAAGCAT
AATATAGCAAGGACTAACCCCTATACCTTCTGCATAATGAATTAACTAGAAATAACTTTGCA
AGGAGAGCCAAAGCTAAGACCCCCGAAACCAGACGAGCTACCTAAGAACAGCTAAAAGAGCA
CACCCGTCTATGAAAAAAAAAAAGCTT

SEQ ID NO. 149: **OC 12-4-1**

GGCTTAAGCTTTTTTTTTTCAAAAATACAAAATAAATTATTTGTAGGCATGGACAATGACAG
CAGTAAACTGNTATTTATTGTCAGCTGAAATCAGTAACTGATGGTTGTAGTGATTTTTTAAA
AACATCACCCAGCATTTTCTTCAGTCATTTTCTTCAAATGACTTCTCTGTAGTTACTGGAGA
GAAATACTGCCTTGAGCTTCCTATCGCCGA

SEQ ID NO. 150: **OG 14-4A**

TCTGTGCTGGGAAACTGGCTAACTGTATGCAAAAAACAGAAACTGGACCCCTTCCTTACACC
TTATACAAAAATTAACTCAAGATGGATTAATTAAGACTTAAACGTAAAACCCAAAACCATAA
AAACCCTAGAAGAAAACCTGGGCAATACCATTCACGACATATACTTGGCAAGGTCC

SEQ ID NO. 160: **OC 15-2-C**

TAAGCTTTTTTTTTTTCGGGTGTGCTCTTTTAGCTGTTCTTAGGTAGCTCGTCTGGTTTCGG

GGGTCTTAGCTTTGGCTCTCCTTGCAAAGTTATTTCTAGTTAATTCATTATGCAGAAGGTAT
AGGGGTTAGTCCTTGCTATATTATGCTTGGTTATAATTTTTCATCTTTCCCTTGCGGTACTA
TATCTATTGCGCCAGGTTTCAATTTCTATCGCCTATACTTTATTTGGGTAAATGGTTTGGCT
AAGGTTGTCTGGTAGTAAGGTGGAGTGGGTTCGGAA

SEQ ID NO. 161: **OA 16-5-II**
AGCCAGCGAAGAAGAAAGGGGAACCAAACAACTCAAATGTGGGCAACCAGCATCTGTCTCAG
GGAAGGAAAGCATGTGAGAGAATTTCTGGTTAATGATTGGGGGTAGAAAAGGCCATTGGAAA
ATAGAACCCCTGGATCCTTTTGGAAAGGTGAGGGTTGGGGTTCTGGGCCTTCTATGTCTCTT
CTGTATCTAAAAAAAAAAAGCTT

SEQ ID NO. 162: **OC 16-4-A**
TAGCCAGCGAAGATAGAAAGGTAGTCCCTGGTCAGTCATTAKTATTGGTAAGAGTTAAAATT
AGCAATATATTTAAATTTCTTTCATTTCATGTACGAGTCTTCCCCCAGCCCTTCACTGGGTG
ATACATGTAAGGATTAGGYGTTAGKGAGACAGCTGTAGTCGYACTCAMCATCTGARCCAAGW
AGATAGTCATCATTTTTCTTTCTCTTGATTYACTTGAAAAAAAAAAAGCTT

SEQ ID NO. 163: **OC 16-7-A**
AAGCTTTTTTTTTTTCAGATGWGWTCATTTTATTATGCTTTTAAAACTTARGTACATGKTAC
ATATATTCATTTTAAATGCCTTGATACAAATAAAAAAGGAAAGCACATATATACAAATAAGA
ATGCCACTATCATGGGATAACTTTGAACCTGCTTAAAGTTTTCTCAATTAACGTATTCACAA
GCTTCAGTACTGTAACTATTCGCTGGCT

SEQ ID NO. 164: **OA 17-4-D**
GACCGCTTGTTAAGAGGAACTGATCTCATATATTTGTATCAGAACTGTATTTTTATGTTATA
TTGTATAGTTTGCTCTCCTGCCCCTCTCCTTAAAACTGAATGGTGCCAATAATTTGATACTA
ATGACTACAAAAAAAGGTAATGCCTCATTTACTAGTATTGTTGTAAAATGAGGAATGTATGT
GAATATTCAGATAACCGAGGATTAACCCTTTAAGTGCTGAATCTTTAAAATTTTAATATATT
TTTTTTTGAGGGAAATCTTTCTAAAATGTATTACGCACTTCCCTGCCTTAGTAAACAGAGTA
TACTGGAGAGTATTTAACCTTTTCTTGATGAGTCATGGCATGATTATAAACATCAGCCCCTT
TTAAAAAAAAAAAGCTT

SEQ ID NO. 165: **OC 17-5-E**
GACCGCTTGTGGATGGAGAAGGGGAGAGCATCTAGGCAGGCAAACAGAAGGGAAGTGGAGTT
AAACCTCTGGCATGAAGTCTGGGAGTAGGGTAGGCTAGGGGGTTTCTTCTATGACACTTGAC
CCTTCCATGCTGGTTCCCAAGCCTATTGGAGGAATGTGGGTGTGGCCGAGGTGATGGCAAGA
AAGGTGCAAGAAAGTGAGCAGTCTGCCTGTGAGTGAGCACAGATGCCGGGGTGTGTGTGTGT
GTGTGTGATTTTCACTGTGGGGTGTGTCTGTGAGAGCTAGCTGCCTTACCCCTCCTTGGCAC
ATAGTAGGCCTTCCATAAATGTTGGATGGATGGATAAATAGATTGGGACCATCAGACCATGA
AAAAAAAAAAGCTT

SEQ ID NO. 166: **OC 17-8-A**
AAGCTTTTTTTTTTTCCAGAAAAAAACAAACATGCAACACTTCGATTTTCAACTTCCAGCAC
CCAAAACTGTGAGAAAATAAATGTCTGTCGTGTAAGCCAACCAGTTTGTGGCATTTTCTTAT
GGCAGCCCTAGAAAAATAACATACAGTTTTCCTCCTATATCTACCTGTCAGTAATGAGAAGG
TTCAAAAGGACACTAGGCTATTGCTTATTAAAAAGAAAACAAACACACAAAAAAACAACTCT
TTCTTGTACTTAGGATATTTTAAGAAGATTATGCAGAACACTTAATTTCTCCCTATTTTCCT
TATACAAGCGGTC

SEQ ID NO. 167: **OC 17-12-A**
AAGCTTTTTTTTTTTCTAAAATTGCAAAAAGGGACGCCACATTGGKGACAGAAAGCCTGGTT

TCACTTCACGGAATAAGCAGTTTGAGATCAATGTCCCAGAAGAGTTTTGACATTCAGGACTT
AAAATAGCAGCAGCAGCAGAGGTAGCTGAAATGGCAAGTAATGAAAATTGCTTTAGTAA
AAATATTTTGGACTGAAGGTATGAGAAACTAAAAGTAGAAACTAGTAAGACACAAAGCATAA
CATGACCAGGAATCTGATACAGTAGTGAACAAGCGGTC

SEQ ID NO. 168: OC 17-12-B

AAGCTTTTTTTTTTCTGATTAAGTTACAAACATTCTCCCTATAGCTAAACTCCGTGACTAG
GCTCCCAGCCTCATGGCCAAGAACAATAAGTTCACCCACTTATCTGGAGTAACCATACTAGA
TTAAAGAAATACAATTCTTTCTTCTAAAGACAATTTCCAGAAAGACCTGCCTTTCCCTATGG
GTACTTGACACTAGGTCCCAGCACAGGCTAATCGCTGTATGGTTTCTTCGAAGATTGGCTTT
TCTCAGTTTCTTTCTCTTTGATACTGTACAAGCGGTC

SEQ ID NO. 169: OC 17-12-D

AAGCTTTTTTTTTTTCTAGAGTGGTTATTGCTCCATCACCTAGGCTTGAGTGCAGKGGTGTG
ATCTTGGCTCACTGCAGCCTCAACCTCCTGGGCCCAAGCAATCCTCCCACCTCAGCCTCTTG
AGTAGCTGGGACCACAGACGTGCACCACGAGACCCAGCTAATTTTTAATTTTTTTTTGTAGA
GGTGGGGGTCTTCCTATGTTGCCCAAGCTGGTCTCAGACTCCTGAGTTCAAGTGATTCTCCC
ACCTAAGCCTCCCAATGTTCTGAGATTACAAGCGGTC

SEQ ID NO. 170: OA 19-5-2

AAGCTTTTTTTTTTAAGATTGTTCTAATTCTGGTTGTAAACTGCTATTTTAAAAAACAAAA
CAAACAGAAAACATCAAAAACACAAAAAGATATTAAAACAGCAAGTCTTTTGTACATCACTG
TAGCATAAGCTGCTTGAGGTTGTCATGCAGAATAGTATCCTTCACGTCACGGAAAACAAGGC
GGATGTTCTCCGTGTTGATAGCAGTGGTGAAGTGGTGGTATAAGGGCTTCTGTTGCTGGTCC
CGACGTTTGAAGC

SEQ ID NO. 171: OC 19-1-1

GCTTCAAACGTCGGATGGGAATTATGTCACCAAACAGGAGCTCAAAGGATTAGATATAGTTA
GAAGAGATTGGTGTGATCTTGCTAAAGACACTGGAAACTTTGTGATTGGCCAGATTCTTTCT
GATATAAACACCAATAGCACCAATCTGGAAGAAGTATTTAAGTTGGGAAACAAGGTAAAAAG
TGAAGTGAATAAGTTGTACAAACTGCTTGAAATAGACATTGATGGGGTTTTCAAGTCTCTGC
TACTGCTGAAAAAAAAAAAGCTTAAGC

SEQ ID NO. 172: OC 19-2-1

GCTTAAGCTTTTTTTTTTTCGCAAACTCATCACTAGACATCGTACTACACGACACGTACTAC
GTTGTAGCCCACTTCCACTATGTCCTATCAATAGGAGCTGTATTTGCCATCATAGGAGGCTT
CATTCACTGATTTCCCCTATTCTCAGGCTACACCCTAGACCAAACCTACGCCAAAATCCATT
TCACTATCATATTCATCGGCGTAAATCTAACTTTCTTCCCACAACACTTTCTCGGCCTATCC
GGAATGCCCCGACGTTTGAAGCC

SEQ ID NO. 173: OA 21-2-2

GGCTTAAGCTTTTTTTTTTTAAGATTGGGNCTAATTCTGGTTGTAAACTGCTATTTTAAAAA
ACAAAACAAACAGAAAACATCAAAAACACAAAAAGATATTAAAACAGCAAGTCTTTTGTACA
TCACTGTAGCATAAGCTGCTTGAGGTTGTCATGCAGAATAGTATCCTTCACGTCACGGAAAA
CAAGGCGGATGTTCTCCGTGTTGATAGCAGTGGTGAAGTGGTGGTATAAGGGCTTCTGTTGC
TGGTCCCGACGTTTGAAGC

SEQ ID NO. 174: OC 24-1-E

AAGCTTTTTTTTTTTCAAGGSTAATCAACAAGCTGAGGGAGTGAAAAAAGAACAAAGAAATC
TGTGACTGCTTGTGATCAATTAGTAAACTTAATTTTTTAGATTAAAATGAAATAATACATGC
AAAGCCCTTGGCACAGTGCCTTGCACATAATACATTTCGGGGTTAAGTTGYGCTAGCTATTC

TGTTATTGATTGNCTTGCCCTTTGTTCTCTGGAAGGTTGGATCTTGCCATTTGGGGATGGCC
AATGGGAAGGCTGAGCAAGACATCAGAGGGTGGGAGGAAAGGAGGTATATTTATTTTCCTTA
CTCCTTCTCTGCTGGGCTTCAATTTTGTCACTAGCTGCATTCCTTTTTATGACCACMACTCC
MGTCC

Among the above-described sequences, Heat Shock Protein 60 (HSP60; SEQ ID NO. 149) was further characterized (**Figure 2, Figure 3, Figure 4**). HSP60, produced primarily in response to pathophysiological stress, is localized to the mitochondrial matrix and facilitates protein folding, translocation and assembly. Northern analyses revealed that the constitutive expression of HSP60 in normal brain is attenuated with increasing age. In stark contrast, HSP60 demonstrated robust expression in GBMs from older patients, inversely correlating with survival. A similar relationship between patient prognosis and the expression of most other major stress-inducible proteins was not observed. Taken together, these results suggest that this selective increase in HSP60 expression is not part of a generalized stress response and that modulation.

6. Sequences identified as being over-expressed in "Young" samples

Using the above-described Using the DDRT-PCR methodology described above, the following sequences were identified as being differentially expressed (ie, over-expressed in "young" tumors as compared to "old" tumors), the following sequences were identified as being differentially expressed in tumor cells taken from patients younger than 45 yrs, of age.

SEQ ID NO. 175: NC 11-3-B
GGCTTAAGCTTTTTTTTTTTCGCAAAATCAGGACAATTCGACAGTCTTTCCCCACTCCTTTC
CCCAAATAGGAACGTAATCTCATATTAAAGGAGAAGCTGAACAAAATGGAATAGATGACTTG
AGAAGGAGAARAGGAGAAAGGAGACCATTACGACTGAGAGAAAATAGTTAATTTTAAGTGAC
ATTTGTGGCACAGGAAGATTGAGAGTTTCATAGKACAAAGAAAGAGGTATCAGAAAAAAGTT
TCCTACCATTACGGYGATTGAAGC

SEQ ID NO. 176: NA 12-2-A
AAGCTTTTTTTTTTTACATAGACGGGTGTGCTCTTTTAGCTGTTCTTAGGTAGCTCGTCTGG
TTTCGGGGGTCTTAGCTTTGGCTCTCCTTGCAAAGTTATTTCTAGTTAATTCATTATGCASA
AGGTATAGGGGTTAGTCCTTGCTATATTATGCTTGGTTATAATTTTTCATCTTTCCCTTGCG
GNACTATATCTATTGCGCCAGGTTTCAATTTCTATCGCCG

SEQ ID NO. 177: NC 15-1-1
AAGCTTTTTTTTTTTTCAGAATANGGGAAAATATATTTTTAAGACAACCTNTTGTGGAAAAGT
TCTGGGACAGTTTTCTCCAAGTGGCTTCTACCCTAAAGTCCCTCTAGCAAAATTTTAGGGTC
TCCACACTCACGACAGATGTCCAGTCCCAAGACATATATCATNTTTTGGCACTTCCCCCAAC
CCCTCTCCAACACGTTCTGAATTAGATTTACCCCAATAACTTTGATTTCTGCGTGTAGATGT
TTCTTCAGGCTATCCTGCCCCTGGTTGGTGGGTTCGGA

SEQ ID NO. 178: NA 15-3-B

AAGCTTTTTTTTTTTAGAGGGTTCTGTGGGCAAATTTAAAGTTGAACTAAGATTCTATCTTG
GACAACCAGCTATCACCAGGCTCGATAGGTTTGTCGCCTCTACCTATAAATCTTCCCACTAT
TTTGCTACATAGACGGGTGTGCTCTTTTAGCTGTTCTTAGGTAGCTCGTCTGGTTTCGGGGG
TCTTAGCTTTGGCTCTCCTTGCAAAGTTATTTCTAGTTAATTCATTATGCAGAAGGTATAGG
GGTTAGTCCTTGCTATATTATGCTTGGTTATAATTTTTCATCTTTCCCTTGCGGTACTATAT
CTATTGCGCCAGGTTTCAATTTCTATCGCTATACTTTATTTGGGTAAATGGTTTGGCTAAGG
TTGTCTGGTAGTAAGGNGGAGTGGGTTCGGAA

SEQ ID NO. 179: NC 21-1-2

GGCTTGTTTCGCTCCAAAGGGTGTATTAATTCTGAATGCTAATCATGAAGACTTRRGTTAGG
ACAACACTTCAAACCAGGAAGTGTGAACTGATTTAGATTATAGCTACACAATTTTCTGTGTG
TTAGATCATGGGGTAGTTTGAGTGTTTTCACATGTATTGCTATAAAATCACAGTGTACCAAG
CTCTGGTTTAATATGCCATTAATACTAATTAATAGAGCTGCTAGTCTCTCTCTGGAAAAAAA
AAAAGCTT

SEQ ID NO. 180: NC 21-1-4R

CAATCGCCGTAATGGTAGGAAACTTTTTTCTGATACCTCTTTCTTTGKRCTATGAAACTCTC
AATCTTCCTGTGCCACAAATGTCACTTAAAATTAACTATTTTCTCTCAGTCGTAATGGTCTC
CTTTCTCCTCTTCTCCTTCTCAAGTCATCTATTCCATTTTGTTCAGCTTCTCCTTTAATATG
AGATTACGTTCCTATTTGGGGAAAGGAGTGGGGAAAGACTGTCGAATTGTCCTGATTTTGCG
AAAAAAAAAAAGCTT

SEQ ID NO. 181: YA 30-2-1

CTGCTGGGACTATGGTACTAAATCCRGNAGATGGGCTGTGTAGCAACTCTCCCAGGGAACAC
ACTAGGGTACTTAGGGAGGTGCTTTGTGGAGCATGTTGAAGCTTTGAGATCTGAGCAGGAGG
CAGTGATGTCCCTGGTCTATTCAGGGAAAGATTTCAGTGTGAAATGGTAAACATCCAATTGA
CAGGATTTAGATTTTGCTTAGTTTTTCTGCTTTTTAATGTTTCTATCCCCCATCTCAGTGTT
TTCTTTATCCATCCCAGTGATGCCTTATTTGAAACTGGGCTTAAACTGCAAAAAGAATGAAG
TTGGATTTAGGAAGCTGTTAGATCATTGAGTGGNGNTGAGAGTGAAGTTCACTAGCAGGGAA
GTTTCCTTGAGCCTAAAATAAAAAGAAAAAATTAAAAAGAATCMYGTTTTTTTAATTWAAAA
AAAAAAGCTTT

SEQ ID NO. 182: OA16-4-A

AAGCTTTTTTTTTTTAAGATAAATGTTGAATTGCAGGAAGAATAACATTTTGGAACAGTAAT
GTGGGATATAAGAAAAAGTCACATAGCTCCAAATTTAGGGTGAGACTTTACATGTCTTAGAA
GACCATTAAGAGGACTTCCAACAAGTAGGGGAGACCAAGTTTCAATTAGGGCAGAAGATAGG
GAAGGAACTCTATAAAGAGACTAAAACTGTGAGGGTTCGCTGGCT

It should be noted that the sequence "GAATTC" at the 5' or 3' ends of the sequences may represent a restriction enzyme site used in characterizing the sequences and does not necessarily constitute part of the differentially expressed sequence.

Table 1

AGE-DEPENDENT GENES ASSOCIATED WITH GLIOMA PATIENT SURVIVAL BY DDRT-PCR ANALYSIS

Normal$^-$, Young$^-$, Old$^+$

- known ESTS (5)
- STAT-induced STAT inhibitor-2
- Fibrillin-15
- NPA6,cri-du-chat
- Ribosomal Protein L7a
- Mitchondrial sequences (3)
- Chaperonin (HSP60)
- Glypican 3 (GPC3)
- CDC42
- Glucosamine-6-Phosphate Deaminase
- Oscillin
- Eph-like Receptor Tyrosine Kinase
- SHOX-b
- Cyclophilin-like Protein, CyP-60
- KIAA0570
- Guanine Nucleotide Binding Protein
- DNA Polymerase (-subunit)
- NOVEL (8)

(Normal$^+$, Young$^+$, Old$^-$)

NOVEL (6)

Example 5

Reverse Northern Screening of RNA

A 4 µl aliquot of the purified cDNA amplicons is then reamplified, using similar conditions as described above, without radioactive isotope, and in the presence of 20 uM dNTP. Following electrophoresis through 1.5-2.0% agarose, the amplicons are purified using QIAquick(R) gel extraction (Qiagen, Inc., Valencia, CA) and reconstituted in a total volume of 40 ul. Duplicate 4 µl aliquots of this gel purified cDNA are reamplified and combined in a total volume of 150 µl for reverse Northern analysis. To this sample, 6 µl of 10N NaOH is added, and the mixture incubated at 4°C for 10 minutes to denature the nucleic acids. The mixture is then diluted 1:1 with 150 µl of 2 M NH_4OAc, 150 µl of which is applied to duplicate nylon membranes presoaked with 1 M NH_4OAc. Wells of the slot-blot apparatus (Schleicher & Schuell, Keene, NY) are washed with 150 µl of 1 M NH_4OAc and filters rinsed in 6X SSC and soaked for 15 minutes in 2X Denhardt's solution, and air dried. The filters are UV-crosslinked in a Stratlinker apparatus (Stratagene, LaJolla, CA), and prehybridized for 2-4 hours at 57°C in 10%

dextran sulfate, 1 M NaCl, 1% SDS, and 50 ug/ml sheared salmon sperm DNA. The radiolabeled probe is prepared by reverse transcription (RT) of 10 ug total RNA from normal fetal astrocytes, or glioma cell line U373MG cells, utilizing the above conditions. Following RT, probe is treated with 20 ug RNase A for 30 minutes at 37°C and purified by Sephadex G50 chromatography. Equivalent amounts of radiolabeled probe (2-3 x 10^6 cpm/ml) are added to the respective blots and hybridized overnight at 57°C. Blots were washed in 2X SSC/1% SDS at 57°C for 30 minutes and autoradiographed for an appropriate time.

The minimal selection criteria for the bands of interest is approximately two-fold greater signal expressed in either tissue, and is qualitatively evaluated by visual inspection of the autoradiographic image.

The amplicons determined to be differentially expressed (either glioblastoma or normal brain tissue specific) are subsequently subcloned into the TA cloning site of the pCR(R)2.1 vector (Invitrogen, Carlsbad, CA) and insert-containing vectors from multiple positive transformants sequenced using an ABI 377 automated fluorescence-based nucleic acid sequencer. All NCBI maintained nucleotide databases (National Center for Biotechnology Information; Bethesda, MD) are searched for homologies using the BLAST (basic local alignment search tool) program. As should be noted by the skilled artisan, use of the TA cloning site occassionally results in the inclusion of a poly-A or poly-T sequence at the 5'- or 3' end, respectively, of the cloned insert. Such sequences are not required to perform the assays described herein.

Following this procedure, the following sequences were found to be overexpressed in tumors of "old" patients as compared to "young" patients. The primers utilized to amplify each of the amplicons is described as well. The sequences are:

OA 3-1-B (SEQ ID NO. 142) NOVEL

GAATTCGCCCTTAGTCAGCCACCATGAACAAAGTGGATCTTGTCTTCTTACATCTATGAAAATAGAGCTTTGAAT
GGTAAGGAGATATGTTTTCTTGGTAACCAATGCAAGATTGATGGGTGGAAACATGATTCAAACTTACACAATTTT
TCTTGCTATTTTTCAAATATGAATCTTACTATATATTCTCGGTGAACATCAGGAGACTATTAAAGAGGTCTGCTG
TTAAATGTAAAAAAAAAAAGCTTAAGGGCGAATTC

OLIGO

	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	19	20	58.26	45.00	6.00	2.00	
CCACCATGAACAAAGTGGAT							
RIGHT PRIMER	205	21	59.15	47.62	5.00	3.00	
CTCCTGATGTTCACCGAGAAT							

SEQUENCE SIZE: 260
PRODUCT SIZE: 187

OA 11-4-1 (SEQ ID NO. 143) STAT-induced STAT inhibitor (STATI2)

GAATTCGGCTTAAGCTTTTTTTTTTTTAGAAATCAGGNGKTTTTTTATTTAATACATTCTAATCAAATAGTAACAG
CAGTAAATAAACACTTTGAAAAACAGGCAGGTATCCCCCTGTATCTGGAAGAAAATTAAGTCAAAGTATTCTACA
CAGTAGAAGGGAGACAACTGTTTATGTCCATGGTTAGACAATTCAAGGACAACTTGGATATTTCTAAAGCCATTT
CCAAAAAATCAATGGCAACAGGTTGGGACACAGCTATTTCAAAGGGTAGAATGCCTATACCTACATTGGTTTTTA
TTAACGGCGATTGAAGCCGAATTC

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	100	20	59.67	55.00	4.00	2.00	
AGGCAGGTATCCCCCTGTAT							
RIGHT PRIMER	266	21	59.99	47.62	6.00	1.00	
TGAAATAGCTGTGTCCCAACC							

SEQUENCE SIZE: 324
PRODUCT SIZE: 167

OA11-5-B/C (SEQ ID NO. 144) NOVEL (fibrillin homology)

GAATTCGCCCTTAAGCTTTTTTTTTTTTAGCGACAGTTGTATTTATTTTTTTTAAGTTACAATA
AAATGCTCTCAAGTCCTTTGAATGTTCCAACAAATTCAAAACTTCATTTTCTGAATGTTTTA
CATAAATGCGAACTACCTGTTCGCATTGGNAACCTGCTGCTGTATTTCATGTCTTAACGGCG
ATTGAAGGGCGAATTC

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	17	22	60.02	40.91	4.00	3.00	
CGCCGTTAAGACATGAAATACA							
RIGHT PRIMER	137	21	58.09	42.86	4.00	3.00	
TGCTCTCAAGTCCTTTGAATG							

SEQUENCE SIZE: 202
PRODUCT SIZE: 121

OC 11-4-C (SEQ ID NO. 147) RIBOSOMAL PROTEIN L7a

GAATTCGCCCTTAAGCTTTTTTTTTTTTCGAAGGAAAATTTGTATTATTTSAATTATTTTTATGKACAGAAAACTC
AACAGTGTACATTTAACCCAGTTTAGKGGCAAGTTCTTTAGCCTTTGCCTTTTCGAGCTTGGCGATACGAGCCAC
AGACTTAGGACCCAGGACACTGCCACCCCAGTGACGGCGATTGAAGGGCGAATTC

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	68	26	57.47	34.62	6.00	2.00	
GAAAACTCAACAGTGTACATTTAACC							
RIGHT PRIMER	173	20	59.72	60.00	3.00	3.00	
GCAGTGTCCTGGGTCCTAAG							

SEQUENCE SIZE: 205
PRODUCT SIZE: 106

OC 12-4-1 (SEQ ID NO. 149) HSP 60

GAATTCGGCTTAAGCTTTTTTTTTTCAAAAATACAAAATAAATTATTTGTAGGCATGGACAATGACAGCAGTAAA
CTGNTATTTATTGTCAGCTGAAATCAGTAACTGATGGTTGTAGTGATTTTTTAAAAACATCACCCAGCATTTTCT
TCAGTCATTTTCTTCAAATGACTTCTCTGTAGTTACTGGAGAGAAATACTGCCTTGAGCTTCCTATCGCCGAAAG
CCGAATTC

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	50	20	59.68	50.00	5.00	2.00	
TAGGCATGGACAATGACAGC							
RIGHT PRIMER	209	22	59.51	50.00	3.00	1.00	
GCTCAAGGCAGTATTTCTCTCC							

SEQUENCE SIZE: 233
PRODUCT SIZE: 160

OC 16-4-A (SEQ ID NO. 162) NOVEL

GAATTCGCCCTTAGCCAGCGAAGATAGAAAGGTAGTCCCTGGTCAGTCATTAKTATTGGTAAGAGTTAAAATTAG
CAATATATTTAAATTTCTTTCATTTCATGTACGAGTCTTCCCCCAGCCCTTCACTGGGTGATACATGTAAGGATT
AGGYGTTAGKGAGACAGCTGTAGTCGYACTCAMCATCTGARCCAAGWAGATAGTCATCATTTTTCTTTCTCTTGA
TTYACTTGAAAAAAAAAAAGCTTARGGGCGAATTC

OLIGO	start	len	tm	gc%	any	3'	seq

LEFT PRIMER	28	22	60.03	50.00	5.00	2.00
AAAGGTAGTCCCTGGTCAGTCA						
RIGHT PRIMER	143	21	57.37	47.62	6.00	0.00
ACATGTATCACCCAGTGAAGG						

SEQUENCE SIZE: 260
PRODUCT SIZE: 116

OA 21-2-2 (SEQ ID NO. 173) Guanine Nucleotide Binding Protein α13 (GNA13)

GAATTCGGCTTAAGCTTTTTTTTTTTAAGATTGGGNCTAATTCTGGTTGTAAACTGCTATTTTAAAAAACAAAAC
AAACAGAAAACATCAAAAACACAAAAAGATATTAAAACAGCAAGTCTTTTGTACATCACTGTAGCATAAGCTGCT
TGAGGTTGTCATGCAGAATAGTATCCTTCACGTCACGGAAAACAAGGCGGATGTTCTCCGTGTTGATAGCAGTGG
TGAAGTGGTGGTATAAGGGCTTCTGTTGCTGGTCCCGACGTTTGAAGCCGAATTC

OLIGO

	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	147	20	61.02	50.00	5.00	2.00	
TGCTTGAGGTTGTCATGCAG							
RIGHT PRIMER	258	20	59.88	50.00	4.00	2.00	
ACCAGCAACAGAAGCCCTTA							

SEQUENCE SIZE: 280
PRODUCT SIZE: 112

OA7-1-B (SEQ ID NO. 184) Guanine Nucleotide Binding Protein α13 (GNA13)

GAATTCGCCCTTCAAACGTCGGGACCAGCAACAGAAGCCCTTATACCACCACTTCACCACTGCTATCAACACGGA
GAACATCCGCCTTGTTTTCCGTGACGTGAAGGATACTATTCTGCATGACAACCTCAAGCAGCTTATGCTACAGTG
ATGTACAAAAGACTTGCTGTTTTAATATCTTTTTGTGTTTTTGATGTTTTCTGTTTGTTTTGTTTTTTAAAATAG
CAGTTTACAACCAGAATTAGAACAATCTTAAAAAAAAAAAAGCTTAAGGGCGAATTC

OLIGO

	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	24	20	59.88	50.00	4.00	2.00	
ACCAGCAACAGAAGCCCTTA							
RIGHT PRIMER	250	26	60.00	34.62	4.00	2.00	
TTGTTCTAATTCTGGTTGTAAACTGC							

SEQUENCE SIZE: 281
PRODUCT SIZE: 227

The following sequences were determined by DDRT-PCR and reverse northern assay to be over-expressed in cancer cells, regardless of the age of the patient from whom the tumor sample is isolated. The sequences and primers utilized to amplify the sequences are shown below:

NC17-10-A,H STM 2 (SEQ ID NO. 68)

GAATTCGCCCTTGACCGCTTGTACTGAAGGGAACAGAGACAGAATGAAATGAAAGAAGGCAG
TTGAACTTCTAGGCTTCTACAGGCAGAAAACAGGCTGATAGAACTGCTCAACTACAGACATG
TTCTACCTTTCTAGAAAAAAAAAAAGCTTAAGGGCGAATTC

OLIGO

	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	17	23	59.94	47.83	4.00	0.00	
GCTTGTACTGAAGGGAACAGAGA							
RIGHT PRIMER	131	25	59.77	44.00	6.00	2.00	
GGTAGAACATGTCTGTAGTTGAGCA							

SEQUENCE SIZE: 165
PRODUCT SIZE: 115

NC17-10-B,C,D NOVEL (SEQ ID NO. 69)

GAATTCGCCCTTGACCGCTTGTTGACAGGATATGGGAGATGGAAAAGGAAAGGATCTGCATC
TAGTGATTGGAAATATAGGAGTGGTGGGGGTTAGTTTCAGATGCCTGTGGGATATTTAATGT
CCTGTGTTGAGTTGGAACTATGAGTTCTACAGAGGGCAAGATTTAGGAGTTGGCACTCCTAA

GTGTCAATACATGTGAATAGGATCGCTTTGGAGGGTGAGAAGAGGTCTGAGAACACTACTAG
GGAACAGTGAAGGAAAAAAAAAAAAGCTTAAGGGCGAATTC

	start	len	tm	gc%	any	3'	seq
OLIGO	34	20	59.87	50.00	2.00	0.00	
LEFT PRIMER							
GGGAGATGGAAAAGGAAAGG							
RIGHT PRIMER	233	20	59.23	55.00	2.00	0.00	
GACCTCTTCTCACCCTCCAA							

SEQUENCE SIZE: 288
PRODUCT SIZE: 200

OA10-1-A (SEQ ID NO. 183) NOVEL

GAATTCGCCCTTGTGATCGCAGTATTCCTTGTATGGAAGTCATCAGATATGCTGTGCAAGTCTTGCTTAATGTAT
CTAAGTATGAGAAAACTACTTCAGCAGTTTATGATGTAGAAAATTGTATAGATATACTATTGGAGCTTTTGCAGA
TATGCCGAGAAAAGCCTGGTAATAAAGTTGCAGACAAAGGCGGAAGCATTTTTACAAAAACTTGTTGTTTGTTGG
CTATTTTACTGAAGACAACAAATAGAGCCTCTGATGTACGAAGTAGGTCCAAAGTTGTTGACCGTATTTACAGTC
TCTACAAACTTACAGCTCATAAACATAAAATGAATACTGAAAGAATACTTTACAAGCAAAAGAAGAATTCTTCTA
TAAGCATTCCTTTTATCCCAGAAACACCTGTAAGGACCAGAATAGTTTCAAGACTTAAGCCAGATTGGGTTTTGA
GAAGAGATAACATGGAAGAAATCACAAATCCCCTGCAAGCTATTCAAATGGTGATGGATACGCTTGGCATTCCTT
ATTAGTAAATGTAAACATTTTCAGTATGTATAGTGNAAAGAAATATTAAAGCCAATCATGAGTACGTAAAAAAAA

	start	len	tm	gc%	any	3'	seq
AAGCTTAAGGGCGAATTC							
OLIGO	187	20	59.72	40.00	3.00	2.00	
LEFT PRIMER							
AAGGCGGAAGCATTTTTACA							
RIGHT PRIMER	489	20	59.93	45.00	4.00	1.00	
CTTGCAGGGGATTTGTGATT							

SEQUENCE SIZE: 618
PRODUCT SIZE: 303

Example 6

Further Selection of Characteristic RNA

Those mRNAs exhibiting differential expression following the reverse Northern screening are chosen for further detailed analysis using clinically relevant tissue and secondary reverse Northern analysis. Individual vector-bound cDNA inserts identified, subcloned and sequenced from the initial screen are linearized with an appropriate restriction enzyme and immobilized on each of six nylon membranes, as described above. The prepared membranes are individually hybridized with radiolabled probes prepared by reverse transcription of 10 ug total RNA from each of three normal brain tissue or three glioblastoma brain tissue samples. Following reverse transcription, the probes are treated with 20 ug RNase A for 30 minutes at 37°C and purified by Sephadex G50 chromatography. Equivalent amounts of radiolabeled probe (1.1-1.2 x 10^6 cpm/ml) are added to the respective blots and hybridized overnight at 57°C. Blots are washed in 2X SSC/1%SDS at 57°C for 30 minutes and analyzed by Phosphor Imaging for 48 hours.

Individual radioactive signals on the blots are quantitated using BioRad Model GS-250 Molecular Imager(R) System and Molecular Analyst(TM)/Macintosh Image Analysis Software

(Version 2.1). One-dimensional profiles are optimized by subtracting image background, as well as pGEM(R) vector control value. An independent Student's t-test is performed comparing the peak heights (in counts) of the three glioblastoma blots, and the three normal brain tissue blots, for each differentially expressed cDNA using SigmaPlot 5.0.

Example 7

Isolation of cDNA Clones Related to Differentially Expressed Sequences

The probes selected above as characteristic signals can then be used to identify gene sequences by screening human cDNA libraries. For example approximately 2 x 10^6 independent clones from a lambda-gt-11 oligo(dT)+random primed human fetal cDNA library (Clontech, Palo Alto, CA) are screened with radiolabeled amplicons from a differentially expressed characteristic signals identified above. Positive plaques are purified by additional screening, and the inserts isolated by subcloning into pGEM(R)7zf(-) vector, sequenced and individually utilized in reverse Northern screening of clinical tissues. The isolated and cloned nucleic acid signals corresponding to the expressed genes of SEQ ID NOS. 1-184 identify the characteristic signals of the invention. Known genes, and the complete nucleic acid sequence for such genes can be obtained from the art, and detection probes designed to specifically identify the expression of such genes in biological samples. In particular, once known, one of ordinary skill in the art can readily identify and prepare hybridization probes which will be suitable for the specific hybridization detection of the desired gene transcript, under a variety of hybridization conditions (*see eg.* Molecular Cloning *supra).* One of skill in the art is able to select and prepare suitable PCR primers for primer specific amplification of the desired gene transcript. Such primers can be designed to utilize the poly-A tail present on such transcripts, so as to specifically identify transcription products. Inserts identified as novel genes can be further cloned and expanded such that a complete nucleic acid sequence is obtained. However, one of skill in the art will be able to use the nucleic acid sequence of the novel inserts identified in SEQ ID NOS. 1-184, to construct suitable hybridization probes, as well as PCR primers for use in specifically identifying transcripts corresponding to the novel gene represented by the insert.

The characteristic signals listed in SEQ ID NOS. 1-184 are not limited to just these signals, as other further characterizing gene transcripts may also be identified and detected in addition to any one or more of the characteristic signals identified in SEQ ID NOS. 1-184.

Example 8

Kit and Screening Assay for Characteristic Nucleic Acids

The characteristic diagnostic signal probes, being selected and identified above, are readily adaptable for use in production of screening assay kits. Such kits can include pre-packaged nucleic acid probes corresponding to at least a fragment of the above identified panel sequences, wherein when the assay kit is designed for hybridization detection, such probes are preferably from 10 to 25 nucleic acids in length.

Diagnostic/detection kits designed for use in hybridization and/or PCR based detection of signals can include appropriate paired primers that are specific for the nucleic acid sequences of the characteristic signals identified above, wherein said primers can be preferably 10 to 20 nucleic acids in length, or as suitable for use in automated detection apparatus. One of ordinary skill in the art would be able to design appropriate probes and PCR primers for the selective identification of the specific characteristic signals as listed in SEQ ID NOS. 1-184, and using corresponding modified nucleic acids as desired. One of skill in the art will be further able to design specific PCR primers which will allow for the identification of actively transcribed genes by using the poly-A tail of such transcripts as a primer target, or as a partially anchored primer target. One of ordinary skill in the art would be able to generate suitable primers, and select appropriate amplification conditions and schemes to practice the present invention, and make modifications thereto. (See for example McPherson et al., PCR Volume 1, Oxford University Press, (1991)).

The detection kits of the invention also provide for sets of primers or hybridization probes which can be used to detect specific nucleic acid signals corresponding to one or more of the characteristic signals identified in SEQ ID NOS. 1-184, where such primers or probes are designed to be used in individual reactions, sequential reactions, or combination reaction, using one or more of the primers or probes in the same reaction mixture.

The diagnostic kits of the invention can further encompass suitable buffers for rehydration of dried probes, or dilution of concentrated probe solutions, or for preparing test samples, as needed to accomplish the designated assays. Diagnostic kits can be further designed to provide only the specific primers needed for PCR amplification and detection of the specific signals.

Detection assays, and the kits incorporating such assays of the invention, need not

provide detection of the entire panel of signals, but may be designed to provide for less than the entire nine signals. The assays and kits can incorporate appropriate positive and negative controls, such as the tublin gene, where such control is proliferation dependent, or proliferation independent in signal production. The assay probes designed for PCR can incorporate the appropriate reaction contols, where the absence of such a signal is an indication that said amplification assay physically failed.

Example 8

Screening and Selection of Anti-Cancer Drugs

Using cell cultures of brain cancer cells, or even individual cancer cells, selection of promising drug candidates, and the evaluation of efficacy of various anti-cancer drugs for treating such cancer can be performed in the laboratory, either manually or using automated apparatus. For example, glioblastoma cells, as described above can be administered various doses of anti-cancer drugs, and screened for expression of specific nucleic acid messages corresponding to the panel shown in SEQ ID NOS. 1-184. Any changes in the expression, or expression levels of any species of nucleic acid from this panel, as compared with normal or control cancer cells, would indicate potential for the therapeutic.

Typical anti-cancer drugs which can be specifically screened include Cytarabine, Fludarabine, 5-Fluorouracil, 6-Mercaptopurine, Methotrexate, 6-Thioguanine, Bleomycin, Dactinomycin, Daunorubicin, Doxorubicin, Idarubicin, Plicamycin, Carmustine, Iomustine, Cyclophosphamide, Ifosfamide, Mechloroethamine, Streptozotocin, Navelbine, Paclitaxel, Vinblastine, Vincristine, Asparaginase, Cisplatin, Carboplatin, Etoposide, Interferons, Procarbazine.

In addition, various sub-types of brain cancer tissues can be screened for their susceptibility to various anti-cancer therapies, by monitoring any change in the characteristic pattern of expressed genes selected from SEQ ID NOS. 1-184, or fragments or complements thereof, as compared with non-malignant cell expression.

Using the present invention, not only can drug candidates can be screened for potential efficacy using standardized malignant cell cultures, biopsy cells may be cultured and used to screen for efficacy as well. While it would be useful to have long term stable cultures of biopsy cells, the assays of the invention can be performed over a short period of time, thus avoiding the

necessity of long term cultures. Thus, the assay of the invention can be performed on specific brain cancer tissue from individual patients, and the potential efficacy of various therapeutics may be tested on those specific cells.

Even if the biopsy sample is not robust enough, or large enough for direct assays of the invention, analysis of the biopsy sample for the characteristic expression of signals, will allow for the selection of a model cancer cell line, which expresses a similar panel of characteristic signals as the biopsy sample. This selected model cell line, and results of therapeutics on the model cell line, may then be used to assess potential therapeutics and treatment.

Example 9

Antisense Inhibition of Gene Expression

The invention encompasses antisense therapeutics which can be used to alter gene expression or RNA translation in targeted cells. Antisense therapy can be accomplished using the identified characteristic nucleic acid insert sequences and genes containing the sequence, the entire gene identified as being characteristic, identified known genes, and suitable fragments of all of these nucleic acids. The design and use of antisense therapeutics is described in the art (see for example Eguchi et al., "Antisense RNA", Ann. Rev. Biochem., 1991, 60:631-52). Even more useful than just the insert fragments, the complete nucleic acid seqence for a novel gene, such as CINN-1, and known genes, allows for the preparation of many more anti-sense nucleic acid therapeutics designed for inhibiting translation of the corresponding protein. All antisense nucleic acids can further incorporate modified backbone structures which give unique functionality to the nucleic acid for use as a therapeutic agent. (See for example Verma & Eckstein, (1998), *Ann. Rev. Biochem.*, 67:99-134).

For example, antisense nucleic acids, either RNA, DNA or PNAs (Protein nucleic acids) can be designed to be complementary for the nucleic acid sequences given as SEQ ID NOS. 1-184, in their entirety, or a selected fragment thereof. In particular, fragments of from 10 to 15 nucleic acids can be designed based on the sequences of the nucleic acids described by SEQ ID NOS. 1-184. An exemplary antisense molecule from which a 10-15-mer oligo may be selected is SEQ ID NO. 184. Smaller or larger fragments may also be designed, however selection for hybridization strength, and half-life duration in use will need to be made using standard criteria of analysis and established practice in the art.

While a preferred form of the invention has been shown in the drawings and described, since variations in the preferred form will be apparent to those skilled in the art, the invention should not be construed as limited to the specific form shown and described, but instead is as set forth in the claims.

CLAIMS

WE CLAIM:

1. A method for ascertaining the propensity of a cell for malignant phenotype said cell being isolated or in a biological sample, said method comprising assaying a cell or biological sample to be tested for a signal indicating the transcription of a nucleic acid transcript, wherein said transcript is from at least one gene selected from the group consisting of nucleic acid sequences identified in SEQ ID NOS. 1-184.

2. A method of claim 1 wherein said nucleic acid is selected from the group consisting of SEQ ID NO. 68, 69, 142, 143, 144, 147, 149, 162, 173, and 183.

3. A method of claim 1 wherein said nucleic acid is selected from the group consisting of SEQ ID NO. 68, 69 and 183.

4. The method of claim 1, wherein said expressed gene is detected by RT-PCR using at least one gene specific amplification primer.

5. The method of claim 1, wherein said expressed gene is detected by nucleic acid hybridization using at least one gene specific probe.

6. The method of claim 5, wherein said assay is *in situ* hybridization.

7. The method of claim 1, wherein a protein encoded by said expressed gene is detected by protein gel assay.

8. The method of claim 1, wherein a protein encoded for by said expressed gene is detected by antibody binding assay.

9. The method of claim 1, wherein said expressed gene is detected by RNase protection assay.

10. The method of claim 1, wherein said gene contains a nucleic acid sequence corresponding to a portion of the nucleic acid sequence selected from the group consisting of SEQ ID NOS. 1-184.

11. A method for ascertaining the suitability of an anti-neoplastic drug candidate for efficacy in treating a malignancy, wherein said method comprises combining said candidate drug with a cell having a cancer phenotype, said cell being isolated or in a biological sample, detecting in said cell or biological sample any change in the expression of at least one of the genes selected

from the group consisting of the nucleic acid sequences of SEQ

12. A method of claim 1 wherein said nucleic acid is selec

of SEQ ID NO. 68, 69, 142, 143, 144, 147, 149, 162, 173, and 183.

13. A method of claim 1 wherein said nucleic acid is selected from the consisting

of SEQ ID NO. 68, 69 and 183.

14. A method of claim 11, wherein said malignant biological sample is a biopsy sa

from a patient to be treated.

15. A method as in claim 14, wherein said malignant biological sample is a cell line.

16. A method as in claim 14, wherein said malignant biological sample is a cell.

17. A therapeutic compound identified in the method of claim 11.

18. A kit comprising hybridization probes specific for at least two nucleic acid sequences selected from the group consisting of the characteristic nucleic acid sequences identified in SEQ ID NOS. 1-184.

19. A kit of claim 18 wherein said nucleic acid is selected from the group consisting of SEQ ID NO. 68, 69, 142, 143, 144, 147, 149, 162, 173, and 183.

20. A kit of claim 18 wherein said nucleic acid is selected from the group consisting of SEQ ID NO. 68, 69 and 183.

21. A kit of claim 18, further comprising suitable reaction buffer components.

22. A kit of claim 18, wherein said probes are suitable for use in PCR amplification of the specific target.

23. A kit of claim 18, wherein said probes are suitable for *in situ* hybridization.

24. A kit comprising probes specific for at least one protein containing an amino acid sequence corresponding to the translation of at least one nucleic acid sequence selected from the group consisting of SEQ ID NOS. 1-184.

25. A kit as in claim 24, where said probe is an antibody, or antigen binding fragment thereof.

26. A kit as in claim 25, where said probe is a polyclonal antibody.

27. A kit as in claim 25, where said probe is a monoclonal antibody.

28. An isolated nucleic acid comprising a nucleic acid sequence selected from the group

NOS. 1-184.

consisting of an expression vector comprising a nucleic acid sequence of claim 28.

A transformed host cell comprising a nucleic acid sequence of claim 28 operably a transcription regulation component.

FIGURE 1

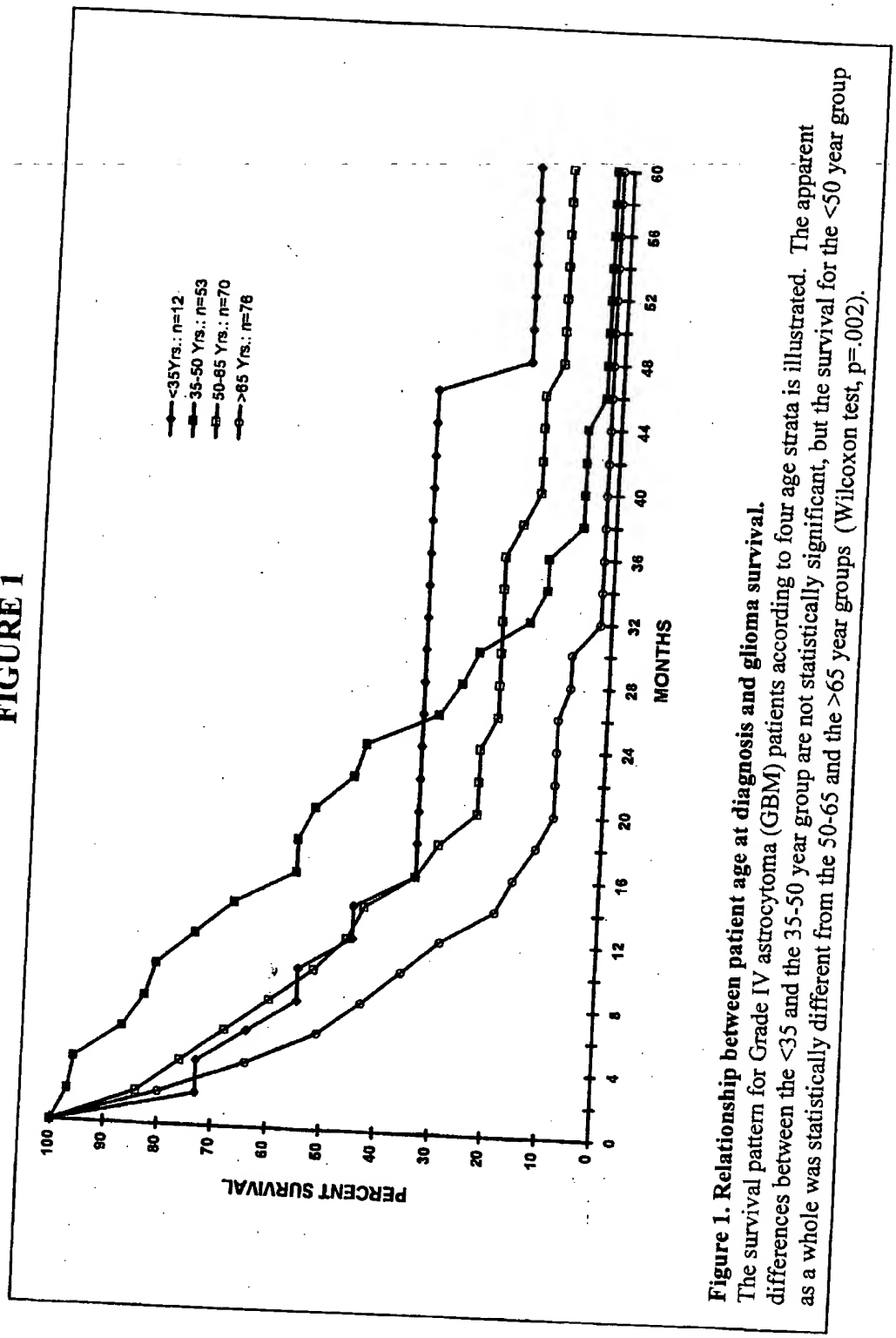

Figure 1. Relationship between patient age at diagnosis and glioma survival.
The survival pattern for Grade IV astrocytoma (GBM) patients according to four age strata is illustrated. The apparent differences between the <35 and the 35-50 year group are not statistically significant, but the survival for the <50 year group as a whole was statistically different from the 50-65 and the >65 year groups (Wilcoxon test, p=.002).

FIGURE 2

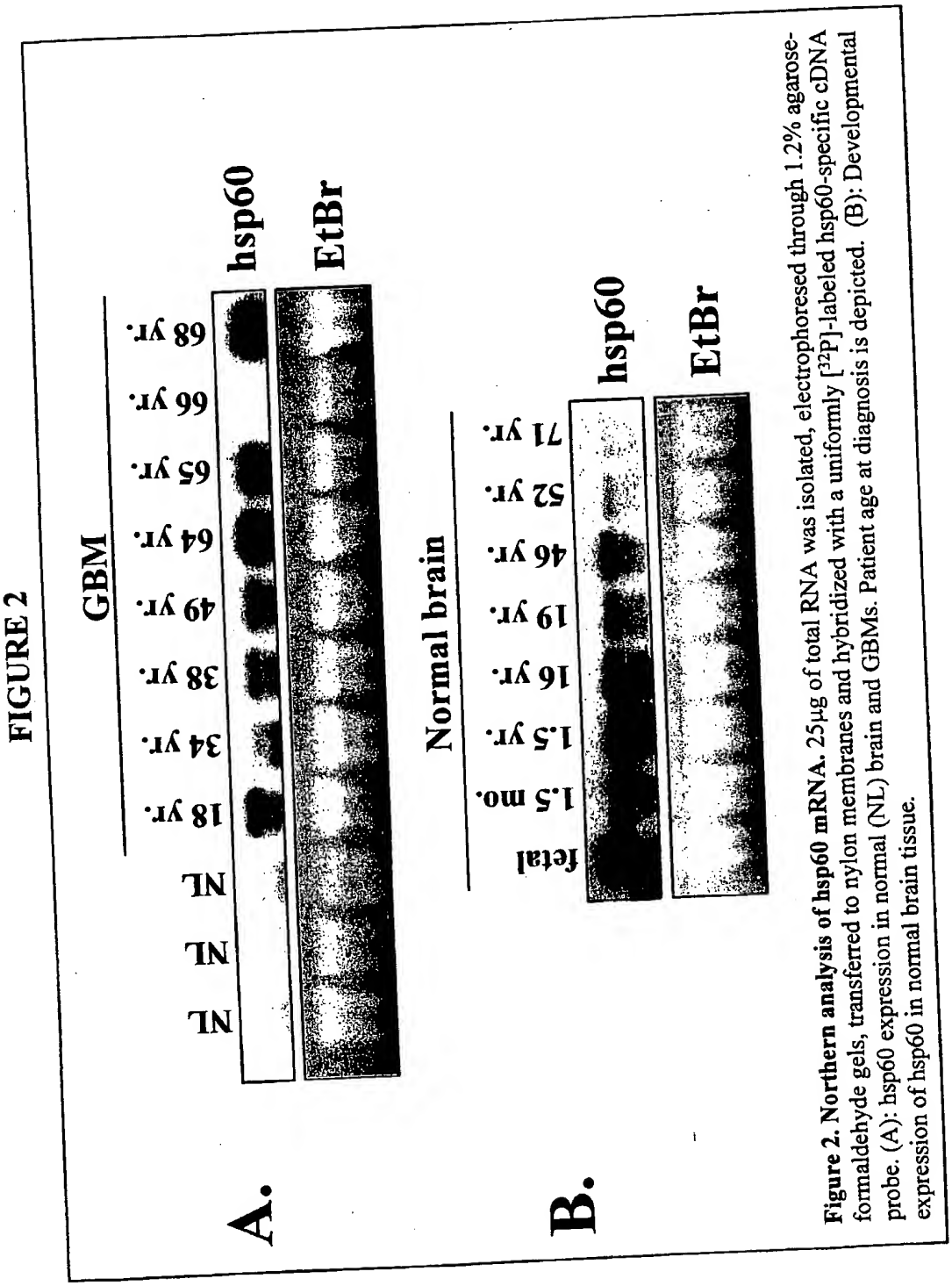

Figure 2. Northern analysis of hsp60 mRNA. 25μg of total RNA was isolated, electrophoresed through 1.2% agarose-formaldehyde gels, transferred to nylon membranes and hybridized with a uniformly [^{32}P]-labeled hsp60-specific cDNA probe. (A): hsp60 expression in normal (NL) brain and GBMs. Patient age at diagnosis is depicted. (B): Developmental expression of hsp60 in normal brain tissue.

FIGURE 3

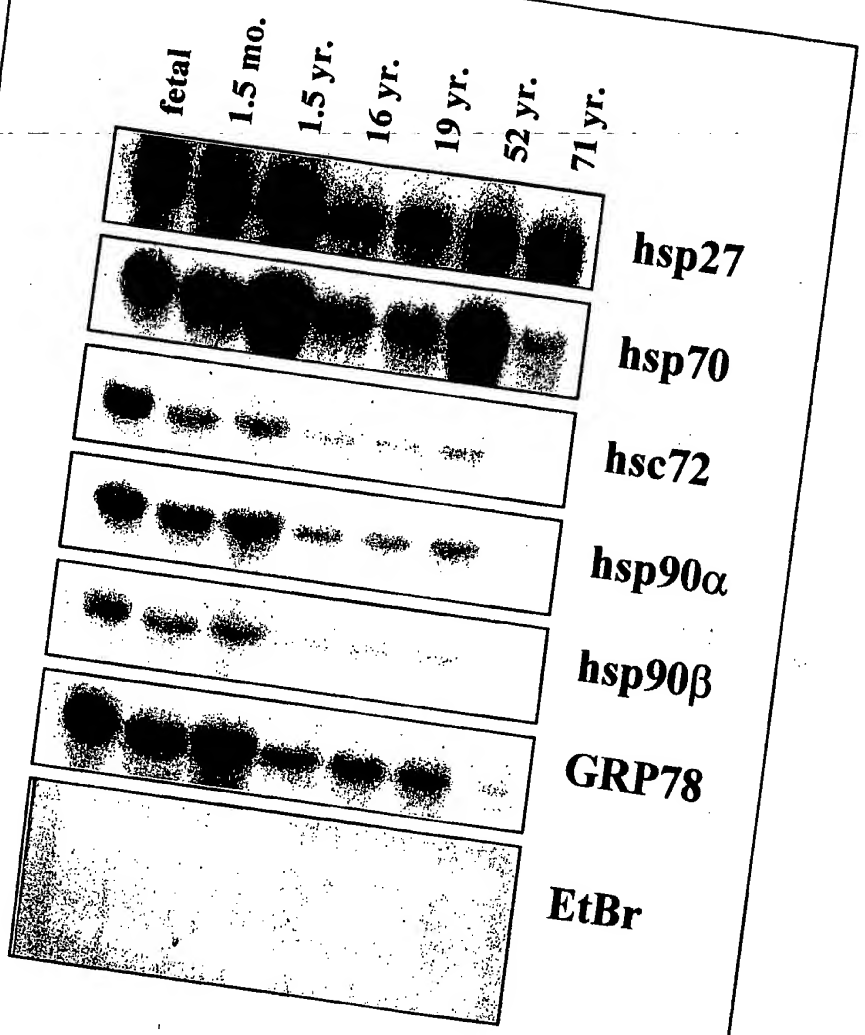

Figure 3. Normal Developmental Expression of Heat Shock Proteins in Human Brain. 25μg of total RNA was isolated, electrophoresed through 1.2% agarose-formaldehyde gels, transferred to nylon membranes and hybridized with a uniformly [^{32}P]-labeled cDNA probes specific for hsp27, hsp70, hsc72, hsp90α, hsp90β, and GRP78.

FIGURE 4

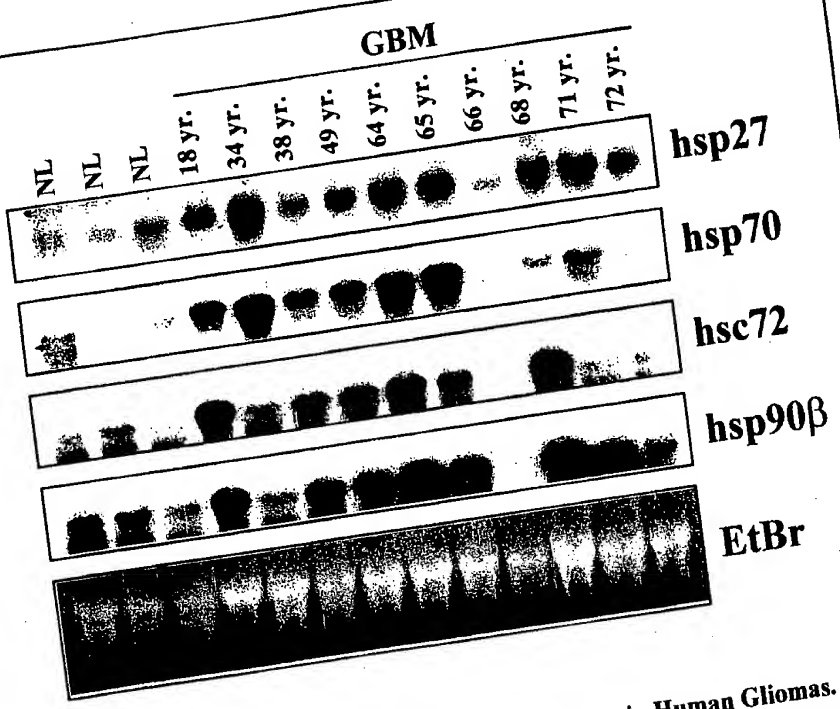

Figure 4. Differential Expression of Heat Shock Proteins in Human Gliomas. 25μg of total RNA was isolated, electrophoresed through 1.2% agarose-formaldehyde gels, transferred to nylon membranes and hybridized with a uniformly [^{32}P]-labeled cDNA probes specific for hsp27, hsp70, hsc72, and hsp90β.